

Chapter 3 Generalized Linear Models

Outline

- 1 3.1 Components of A Generalized Linear Model
- 2 3.2 Generalized Linear Models for Binary Data
- 3 3.3 Generalized Linear Models for Count Data
- 4 3.4 Statistical Inference and Model Checking
- 5 3.5 Fitting Generalized Linear Models
- 6 Homework 3

Outline

- 1 3.1 Components of A Generalized Linear Model
- 2 3.2 Generalized Linear Models for Binary Data
- 3 3.3 Generalized Linear Models for Count Data
- 4 3.4 Statistical Inference and Model Checking
- 5 3.5 Fitting Generalized Linear Models
- 6 Homework 3

All generalized linear models have three components: The *random component* identifies the response variable Y and assumes a probability distribution for it. The *systematic component* specifies the explanatory variables for the model. The *link function* specifies a function of the expected value (mean) of Y , which the GLM relates to the explanatory variables through a prediction equation having linear form.

3.1.1 Random Component

Random Component

The *random component* of a GLM identifies the response variable Y and selects a probability distribution for it. Denote the observations on Y by (Y_1, \dots, Y_n) . Standard GLMs treat Y_1, \dots, Y_n as independent.

3.1.1 Random Component

Remark

- In many applications, the observations on Y are binary, such as “success” or “failure”. More generally, each Y_i might be the number of successes out of a certain fixed number of trials. In either case, we assume a *binomial distribution* for Y .
- In some applications, each observation is a count. We might then assume a distribution for Y that applies to all the nonnegative integers, such as the *Poisson* or *negative binomial*.
- If each observation is continuous, such as a subject’s weight in a dietary study, we might assume a *normal distribution* for Y .

3.1.2 Systematic Component

Systematic Component

The *systematic component* of a GLM specifies the explanatory variables. These enter linearly as predictors on the right-hand side of the model equation. That is, the systematic component specifies the variables that are the $\{x_j\}$ in the formula

$$\alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

This linear combination of the explanatory variables is called the *linear predictor*.

Some $\{x_j\}$ can be based on others in the model. For example, perhaps $x_3 = x_1 x_2$, to allow interaction between x_1 and x_2 in their effects on Y , or perhaps $x_3 = x_1^2$, to allow a curvilinear effect of x_1 .

3.1.3 Link Function

Link Function

Denote the expected value of Y , the mean of its probability distribution, by $\mu = E(Y)$. The third component of a GLM, the *link function*, specifies a function $g(\cdot)$ that relates μ to the linear predictor as

$$g(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

The function $g(\cdot)$, the link function, connects the random and systematic components.

3.1.3 Link Function

Identity Link

The simplest link function is $g(\mu) = \mu$. This models the mean directly and is called the *identity link*. It specifies a linear model for the mean response,

$$\mu = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k.$$

This is the form of ordinary regression models for continuous responses.

Loglinear Model

A GLM that uses the log link is called a *loglinear model*. It has form

$$\log(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k.$$

3.1.3 Link Function

Logistic Regression

The link function $g(\mu) = \log[\mu/(1-\mu)]$ models the log of an odds. It is appropriate when μ is between 0 and 1, such as a probability. This is called the *logit link*. A GLM that uses the logit link is called a *logistic regression model*.

Natural Parameter

Each potential probability distribution for Y has one special function of the mean that is called its *natural parameter*. For the normal distribution, it is the mean itself. For the binomial, the natural parameter is the logit of the success probability. The link function that uses the natural parameter as $g(\mu)$ in the GLM is called the *canonical link*.

3.1.4 Normal GLM

A GLM generalizes ordinary regression models in two ways:

- It allows Y to have a distribution other than the normal.
- It allows modeling some function of the mean. Both generalizations are important for categorical data.

3.1.4 Normal GLM

- Ordinary regression models for continuous responses are special cases of GLMs. They assume a normal distribution for Y and model its mean directly, using the identity link function, $g(\mu) = \mu$.
- The GLM choice of link function is separate from the choice of random component. It is not chosen to produce normality or stabilize the variance.

Outline

- 1 3.1 Components of A Generalized Linear Model
- 2 3.2 Generalized Linear Models for Binary Data
- 3 3.3 Generalized Linear Models for Count Data
- 4 3.4 Statistical Inference and Model Checking
- 5 3.5 Fitting Generalized Linear Models
- 6 Homework 3

This section introduces GLMs for binary responses. Although GLMs can have multiple explanatory variables, for simplicity we introduce them using a single x . The value of π can vary as the value of x changes, and we replace π by $\pi(x)$ when we want to describe its dependence on that value.

3.2.1 Linear Probability Model

Linear Probability Model

In ordinary regression, $\mu = E(Y)$ is a linear function of x . For a binary response, an analogous model is

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

This is called a *linear probability model*, because the probability of success changes linearly in x . The parameter β represents the change in the probability per unit change in x . This model is a GLM with binomial random component and identity link function.

3.2.2 Example: Snoring and Heart Disease

Table 3.1 Relationship Between Snoring and Heart Disease

Snoring	Heart Disease		Proportion Yes	Linear Fit	Logit Fit	Probit Fit
	Yes	No				
Never	24	1355	0.017	0.017	0.021	0.020
Occasional	35	603	0.055	0.057	0.044	0.046
Nearly every night	21	192	0.099	0.096	0.093	0.095
Every night	30	224	0.118	0.116	0.132	0.131

Note: Model fits refer to proportion of “yes” responses.

Source: P. G. Norton and E. V. Dunn, *Br. Med. J.*, 291: 630-632, 1985, published by BMJ Publishing Group.

3.2.2 Example: Snoring and Heart Disease

Example

Table 3.1 is based on an epidemiological survey of 2484 subjects to investigate snoring as a possible risk factor for heart disease. The subjects were classified according to their snoring level, as reported by their spouses. The linear probability model states that the probability of heart disease $\pi(x)$ is a linear function of the snoring level x . We treat the rows of the table as independent binomial samples with probabilities $\pi(x)$. We use scores (0, 2, 4, 5) for $x = \text{snoring level}$, treating the last two snoring categories as closer than the other adjacent pairs.

3.2.2 Example: Snoring and Heart Disease

$\hat{\pi} = 0.0172 + 0.0198x$. The estimated values of $E(Y)$ for a GLM are called *fitted values*.

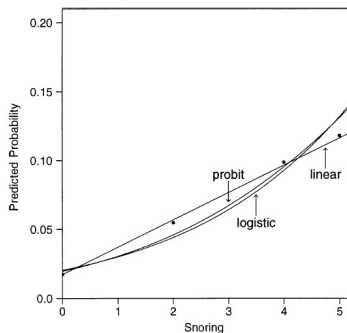


Figure 3.1. Fit of models for snoring and heart disease data.

3.2.2 Example: Snoring and Heart Disease

Suppose we had chosen snoring-level scores with different relative spacings than the scores 0, 2, 4, 5. Examples are 0, 2, 4, 4.5 or 0, 1, 2, 3. Then the fitted values would change somewhat. They would not change if the relative spacings between scores were the same.

If we entered the data as 2484 binary observations of 0 or 1 and fitted the model using ordinary least squares rather than ML, we would obtain $\hat{\pi} = 0.0169 + 0.0200x$.

3.2.3 Logistic Regression Model

Relationships between $\pi(x)$ and x are usually nonlinear rather than linear. The S-shaped curves displayed in Figure 3.2 are often realistic shapes for the relationship.

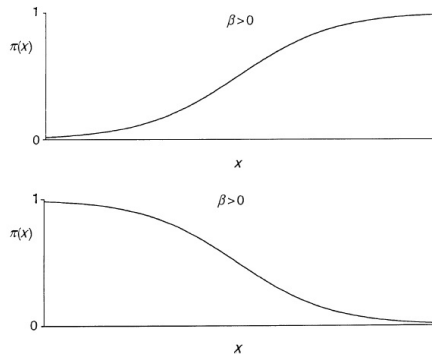


Figure 3.2. Logistic regression functions.

3.2.3 Logistic Regression Model

Logistic Regression

In practice, $\pi(x)$ often either increases continuously or decreases continuously as x increases. The most important mathematical function with this shape has formula using the exponential function

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

This is called the *logistic regression function*. The corresponding logistic regression model form is

$$\log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x.$$

3.2.3 Logistic Regression Model

Logistic Regression

- The logistic regression model is a special case of a GLM. The random component for the (success, failure) outcomes has a binomial distribution.
- The link function is the logit function $\log[\pi/(1 - \pi)]$ of π , symbolized by “logit(π)”.
- Logistic regression models are often called logit models.

3.2.3 Logistic Regression Model

Remark

π is restricted to the 0-1 range, the logit can be any real number. The real numbers are also the potential range for linear predictors (such as $\alpha + \beta x$) that form the systematic component of a GLM, so this model does not have the structural problem that the linear probability model has.

3.2.3 Logistic Regression Model

- The parameter β determines the rate of increase or decrease of the curve.
- When $\beta > 0$, $\pi(x)$ increases as x increases, as in Figure 3.2(a).
- When $\beta < 0$, $\pi(x)$ decreases as x increases, as in Figure 3.2(b).
- When $\beta = 0$, the curve flattens to a horizontal straight line.
- The magnitude of β determines how fast the curve increases or decreases. As $|\beta|$ increases, the curve has a steeper rate of change.

For the logistic regression model for the snoring and heart disease data in Table 3.1, software reports the ML fit, $\text{logit}[\hat{\pi}(x)] = -3.87 + 0.40x$.

3.2.4 Probit Regression Model

Probit Model

Another model that has the S-shaped curves of Figure 3.2 is called the *probit model*. The link function for the model, called the *probit link*, transforms probabilities to z-scores from the standard normal distribution. The probit model has expression

$$\text{probit}[\pi(x)] = \alpha + \beta x.$$

3.2.4 Probit Regression Model

The probit link function applied to $\pi(x)$ gives the standard normal z-score at which the left-tail probability equals $\pi(x)$.

Probit Model

- $\text{probit}(0.05) = -1.645$, because 5% of the standard normal distribution falls below -1.645.
- $\text{probit}(0.50) = 0$, $\text{probit}(0.95) = 1.645$, and $\text{probit}(0.975) = 1.96$.

3.2.4 Probit Regression Model

Example

For the snoring and heart disease data with scores $\{0, 2, 4, 5\}$ for snoring level, software reports that the ML fit of the probit model is

$$\text{probit}[\hat{\pi}(x)] = -2.061 + 0.188x.$$

At snoring level $x = 0$, the probit equals $-2.061 + 0.188(0) = -2.06$. The fitted probability $\hat{\pi}(0)$ is the left-tail probability for the standard normal distribution at $z = -2.06$, which equals 0.020. At snoring level $x = 5$, the probit equals $-2.061 + 0.188(5) = -1.12$, which corresponds to a fitted probability of 0.131.

3.2.5 Binary Regression and Cumulative Distribution Functions

- A monotone regression curve such as the first one in Figure has the shape of a cumulative distribution function (cdf) for a continuous random variable. This suggests a model for a binary response having form $\pi(x) = F(x)$ for some cdf.
- Using an entire class of location-scale cdf's, such as normal cdf's with their variety of means and variances, permits the curve $\pi(x) = F(x)$ to have flexibility in the rate of increase and in the location where most of that increase occurs. Let $\Phi(\cdot)$ denote the standard cdf of the class, such as the $N(0, 1)$ cdf.

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

3.2.5 Binary Regression and Cumulative Distribution Functions

- Shapes of different cdf's in the class occur as α and β vary. Replacing x by βx permits the curve to increase at a different rate than the standard cdf (or even to decrease if $\beta < 0$); varying α moves the curve to the left or right.
- When $\Phi()$ is strictly increasing over the entire real line, its inverse function $\Phi()^{-1}$ exists, equivalently

$$\Phi^{-1}[\pi(x)] = \alpha + \beta x.$$

3.2.5 Binary Regression and Cumulative Distribution Functions

- For this class of cdf shapes, the link function for the GLM is Φ^{-1} . The link function maps the $(0, 1)$ range of probabilities onto $(-\infty, \infty)$, the range of linear predictors. The curve has the shape of a normal cdf when Φ is the standard normal cdf. It is then called the probit model. The parameters of the normal distribution relate to the parameters in the probit model by mean $\mu = -\alpha/\beta$ and standard deviation $\sigma = 1/|\beta|$.
- For the snoring and heart disease data, the probit fit corresponds to a normal cdf having mean $-\hat{\alpha}/\hat{\beta} = 2.061/0.188 = 11.0$, and standard deviation $1/|\hat{\beta}| = 1/0.188 = 5.3$.

3.2.5 Binary Regression and Cumulative Distribution Functions

Application

A typical experiment exposes animals (typically insects or mice) to various dosages of some potentially toxic substance. For each subject, the response is whether it dies. It is natural to assume a tolerance distribution for subjects' responses. For example, each insect may have a certain tolerance to an insecticide, such that it dies if the dosage level exceeds its tolerance and survives if the dosage level is less than its tolerance. Tolerances would vary among insects. If a cdf F describes the distribution of tolerances, then the model for the probability $\pi(x)$ of death at dosage level x the above form.

Outline

- 1 3.1 Components of A Generalized Linear Model
- 2 3.2 Generalized Linear Models for Binary Data
- 3 3.3 Generalized Linear Models for Count Data
- 4 3.4 Statistical Inference and Model Checking
- 5 3.5 Fitting Generalized Linear Models
- 6 Homework 3

Many discrete response variables have counts as possible outcomes. Examples are Y = number of parties attended in the past month, for a sample of students, etc. The simplest GLMs for count data assume a Poisson distribution for the random component.

Poisson Distribution

- $Y = 0, 1, \dots, \mu > 0, E(Y) = Var(Y) = \mu$
- Probability
- As the mean increases the skew decreases and the distribution becomes more bell-shaped. Figure 3.3 shows Poisson distributions with means 2 and 6.

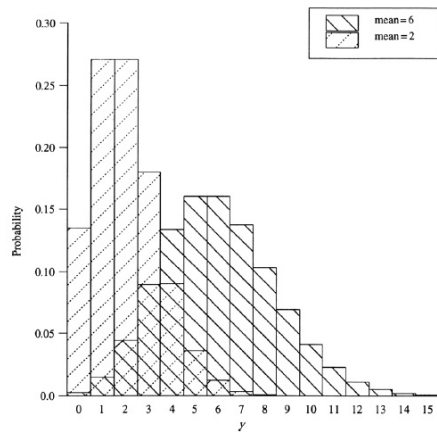


Figure 3.3. Poisson distributions having means 2 and 6.

3.3.1 Poisson Regression

Poisson Loglinear Model

A Poisson loglinear model is a GLM that assumes a Poisson distribution for Y and uses the log link function. For a single explanatory variable x , the Poisson loglinear model has form

$$\log \mu = \alpha + \beta x.$$

The mean satisfies the exponential relationship

$$\mu = \exp(\alpha + \beta x) = e^{\alpha} (e^{\beta})^x.$$

3.3.1 Poisson Regression

Note

- A one-unit increase in x has a multiplicative impact of e^β on μ : The mean of Y at $x + 1$ equals the mean of Y at x multiplied by e^β .
- $\beta = 0$, then $e^\beta = 1$ and the multiplicative factor is 1. Then, the mean of Y does not change as x changes.
- $\beta > 0$, then $e^\beta > 1$, and the mean of Y increases as x increases.
- $\beta < 0$, the mean decreases as x increases.

3.3.2 Example: Female Horseshoe Crabs and their Satellites

Example

Table 3.2 comes from a study of nesting horseshoe crabs (J. Brockmann, *Ethology*, **102**: 1-21, 1996). Each female horseshoe crab in the study had a male crab attached to her in her nest. The study investigated factors that affect whether the female crab had any other males, called *satellites*, residing nearby her. The response outcome for each female crab is her number of satellites. An explanatory variable thought possibly to affect this was the female crab's shell width, which is a summary of her size. In the sample, this shell width had a mean of 26.3 cm and a standard deviation of 2.1 cm.

3.3.2 Example: Female Horseshoe Crabs and their Satellites

Figure 3.4 plots the response counts against width (in centimeters), with numbered symbols indicating the number of observations at each point. To obtain a clearer picture of any trend, we grouped the female crabs into a set of width categories, (≤ 23.25 , 23.25-24.25, 24.25-25.25, 25.25-26.25, 26.25-27.25, 27.25-28.25, 28.25-29.25, >30.25), and calculated the sample mean number of satellites in each category. Figure 3.5 plots these sample means against the sample mean width for the female crabs in each category.

3.3.2 Example: Female Horseshoe Crabs and their Satellites

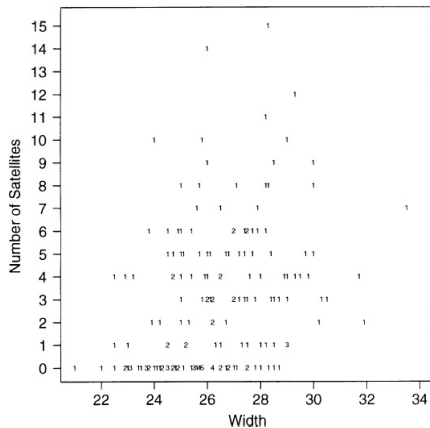


Figure 3.4. Number of satellites by width of female crab.

3.3.2 Example: Female Horseshoe Crabs and their Satellites

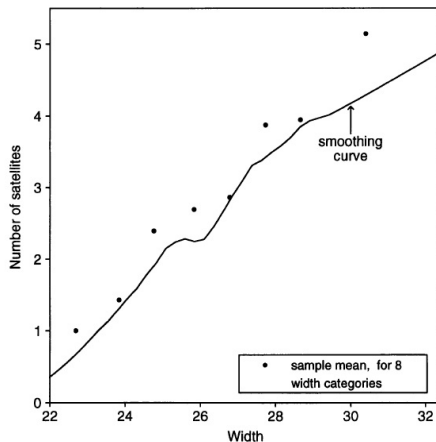


Figure 3.5. Smoothings of horseshoe crab counts.

3.3.2 Example: Female Horseshoe Crabs and their Satellites

Example

Let μ denote the expected number of satellites for a female crab, and let x denote her width. the ML fit of the Poisson loglinear model is

$$\log \hat{\mu} = \hat{\alpha} + \hat{\beta}x = -3.305 + 0.164x.$$

The fitted value at the mean width of $x = 26.3$ is

$$\hat{\mu} = \exp(\hat{\alpha} + \hat{\beta}x) = 2.7.$$

3.3.2 Example: Female Horseshoe Crabs and their Satellites

Example

The Poisson regression model with identity link function has ML fit

$$\hat{\mu} = -11.53 + 0.550x.$$

On average, a 2 cm increase in width corresponds to about an extra satellite.

3.3.2 Example: Female Horseshoe Crabs and their Satellites

Figure 3.6 plots the fitted number of satellites against width, for the models with log link and with identity link.

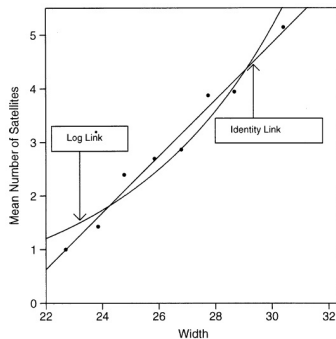


Figure 3.6. Estimated mean number of satellites for log and identity links.

3.3.3 Overdispersion: Greater Variability than Expected

Count data often vary more than we would expect if the response distribution truly were Poisson. The phenomenon of the data having greater variability than expected for a GLM is called *overdispersion*.

Table 3.3 Sample Mean and Variance of Number of Satellites

width	No. Cases	No. Satellites	Sample Mean	Sample Variance
≤ 23.25	14	14	1.00	2.77
23.25-24.25	14	20	1.43	8.88
24.25-25.25	28	67	2.39	6.54
25.25-26.25	39	105	2.69	11.38
26.25-27.25	22	63	2.86	6.88
27.25-28.25	24	93	3.87	8.81
28.25-29.25	18	71	3.94	16.88
> 29.25	14	72	5.14	8.29

3.3.3 Overdispersion: Greater Variability than Expected

Remark

- A common cause of overdispersion is heterogeneity among subjects. If the variance equals the mean when all relevant variables are controlled, it exceeds the mean when only a subset of those variables is controlled.
- Overdispersion is not an issue in ordinary regression models assuming normally distributed Y , because the normal has a separate parameter from the mean (i.e., the variance, σ^2) to describe variability. For Poisson distributions, however, the variance equals the mean. Overdispersion is common in applying Poisson GLMs to counts.

3.3.4 Negative Binomial Regression

Negative Binomial Distribution

The negative binomial distribution is a discrete probability distribution of the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of failures (denoted r) occurs. The negative binomial distribution has

$$E(Y) = \mu, \quad Var(Y) = \mu + D\mu^2.$$

The index D , which is nonnegative, is called a *dispersion parameter*.

3.3.4 Negative Binomial Regression

Negative Binomial Regression

Negative binomial GLMs for counts express μ in terms of explanatory variables. Most common is the log link, as in Poisson loglinear models, but sometimes the identity link is adequate. It is common to assume that the dispersion parameter D takes the same value at all predictor values, much as regression models for a normal response take the variance parameter to be constant. The model can be fitted with GLM software.

3.3.4 Negative Binomial Regression

Example

Consider the horseshoe crab dataset above on Y = number of satellites, with x = shell width as predictor. The Poisson GLM with log link has

$$\log(\hat{\mu}) = -3.30 + 0.164x$$

with $SE = 0.020$ for $\hat{\beta}$. The negative binomial GLM has

$$\log(\hat{\mu}) = -4.05 + 0.192x$$

with $SE = 0.048$ for $\hat{\beta}$.

3.3.5 Count Regression for Rate Data

Rate Data

When events occur over time, space, or some other index of size, models can focus on the rate at which the events occur. For example, in analyzing numbers of murders in 2006 for a sample of cities, we could form a rate for each city by dividing the number of murders by the city's population size. A model might describe how the rate depends on explanatory variables such as the city's unemployment rate, median income, and percentage of residents having completed high school.

3.3.5 Count Regression for Rate Data

Model

A loglinear model for the expected rate has form

$$\log(\mu/t) = \alpha + \beta x.$$

$$\log \mu - \log t = \alpha + \beta x.$$

The adjustment term, $-\log t$, to the log of the mean is called an *offset*. Standard GLM software can fit a model having an offset term.

$$\mu = t \exp(\alpha + \beta x)$$

3.3.6 Example: British Train Accidents over Time

Table 3.4. Collisions Involving Trains in Great Britain

Year	Train-km	Train Collisions	Train-road Collisions	Year	Train-km	Train Collisions	Train-road Collisions
2003	518	0	3	1988	443	2	4
2002	516	1	3	1987	397	1	6
2001	508	0	4	1986	414	2	13
2000	503	1	3	1985	418	0	5
1999	505	1	2	1984	389	5	3
1998	487	0	4	1983	401	2	7
1997	463	1	1	1982	372	2	3
1996	437	2	2	1981	417	2	2
1995	423	1	2	1980	430	2	2
1994	415	2	4	1979	426	3	3
1993	425	0	4	1978	430	2	4
1992	430	1	4	1977	425	1	8
1991	439	2	6	1976	426	2	12
1990	431	1	2	1975	436	5	2
1989	436	4	4				

Source: British Department of Transport.

3.3.6 Example: British Train Accidents over Time

Table 3.4 lists the number of two types of train-related accidents in the UK between 1975 and 2003: (1) accidents involving trains alone (collisions, derailments, and overruns); and (2) collisions between trains and road vehicles. Here we consider only the second count. The table also shows the annual number of train-kilometers, which is a measure of railway activity indicating the millions of kilometers traveled by trains during the year.

3.3.6 Example: British Train Accidents over Time

Let μ denote the expected value of $Y =$ annual number of collisions between trains and road vehicles, for t million kilometers of train travel. During the past decade, rail travel has become increasingly privatized in the UK, and some people have expressed fears that accidents have become more likely. To allow for a trend over time, we consider $x =$ number of years since 1975. Assuming a Poisson distribution for Y , we get the ML fit

3.3.6 Example: British Train Accidents over Time

$$\log(\hat{\mu}) - \log(t) = -4.21 - 0.0329x.$$

Similar results occur with a negative binomial model. The ML fit is

$$\log(\hat{\mu}) - \log(t) = -4.20 - 0.0337x.$$

Outline

- 1 3.1 Components of A Generalized Linear Model
- 2 3.2 Generalized Linear Models for Binary Data
- 3 3.3 Generalized Linear Models for Count Data
- 4 3.4 Statistical Inference and Model Checking**
- 5 3.5 Fitting Generalized Linear Models
- 6 Homework 3

3.4.1 Inference about Model Parameters

Wald

A Wald 95% confidence interval for a model parameter β equals $\hat{\beta} \pm 1.96(SE)$, where SE is the standard error of $\hat{\beta}$. To test $H_0 : \beta = 0$, the Wald test statistic

$$z = \hat{\beta}/SE$$

has an approximate standard normal distribution when $\beta = 0$. Equivalently, z^2 has an approximate chi-squared distribution with $df = 1$.

3.4.1 Inference about Model Parameters

LR

For the likelihood-ratio approach, denote the maximized value of the likelihood function by l_0 under $H_0 : \beta = 0$ and by l_1 when β need not equal 0. The *likelihood ratio* test statistic equals

$$-2 \log(l_0/l_1) = -2[\log(l_0) - \log(l_1)] = -2(L_0 - L_1)$$

where L_0 and L_1 denote the maximized log-likelihood functions. Under $H_0 : \beta = 0$, this test statistic has a large-sample chi-squared distribution with $df = 1$.

3.4.1 Inference about Model Parameters

LR

The 95% confidence interval consists of all β_0 values for which the P-value exceeds 0.05 in the likelihood-ratio test of $H_0 : \beta = \beta_0$. For small n , this is preferable to the Wald interval.

3.4.2 Example: Snoring and Heart Disease Revisited

Example

Section 3.2.2 used a linear probability model to describe the probability of heart disease $\pi(x)$ as a linear function of snoring level x for data in Table 3.1. The ML model fit is $\hat{\pi} = 0.0172 + 0.0198x$, where the snoring effect $\hat{\beta} = 0.0198$, $SE = 0.0028$.

3.4.2 Example: Snoring and Heart Disease Revisited

- The Wald test of $H_0 : \beta = 0$ against $H_a : \beta \neq 0$ treats $z = \hat{\beta}/SE = 7.1$ as standard normal, or $z^2 = 50.0$ as chi-squared with $df = 1$. This provides extremely strong evidence of a positive snoring effect on the probability of heart disease ($P < 0.0001$).
- We obtain similar strong evidence from a likelihood-ratio test comparing this model to the simpler one having $\beta = 0$. That chi-squared statistic equals $-2(L_0 - L_1) = 65.8$ with $df = 1$ ($P < 0.0001$).
- The likelihood-ratio 95% confidence interval for β is (0.0145, 0.0255).

3.4.3 The Deviance

L_M , the maximized log-likelihood value for a model M of interest, L_S , denote the maximized log-likelihood value for the most complex model possible. This model has a separate parameter for each observation, and it provides a perfect fit to the data. The model is said to be *saturated*.

3.4.3 The Deviance

Deviance

Because the saturated model has additional parameters, $L_S \geq L_M$ for a simpler model M . The *deviance* of a GLM is defined as

$$\text{Deviance} = -2[L_M - L_S].$$

The deviance is the likelihood-ratio statistic for comparing model M to the saturated model. It is a test statistic for the hypothesis that all parameters that are in the saturated model but not in model M equal zero.

3.4.4 Model Comparison Using the Deviance

Analysis of Deviance

Consider two models, denoted by M_0 and M_1 , such that M_0 is a special case of M_1 . For normal-response models, the F -test comparison of the models decomposes a sum of squares representing the variability in the data. This *analysis of variance* for decomposing variability generalizes to an *analysis of deviance* for GLMs. Given that the more complex model holds, the likelihood-ratio statistic for testing that the simpler model holds is $-2[L_0 - L_1]$. Since

$$-2[L_0 - L_1] = -2[L_0 - L_S] - \{-2[L_1 - L_S]\} = \text{Deviance}_0 - \text{Deviance}_1$$

we can compare the models by comparing their deviances.

3.4.5 Residuals Comparing Observations to the Model Fit

Pearson Residual

The *Pearson residual* is a standardized difference

$$\text{Pearson residual} = e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{\text{Var}}(y_i)}}.$$

When the GLM is the model corresponding to independence for cells in a two-way contingency table, this is the Pearson chi-squared statistic X^2 for testing independence. Components of the deviance, called *deviance residuals*, are alternative measures of lack of fit.

3.4.5 Residuals Comparing Observations to the Model Fit

Standardized Residual

The *standardized residual* takes $(y_i - \hat{\mu}_i)$ and divides it by its estimated standard error, that is

$$\frac{y_i - \hat{\mu}_i}{SE}$$

It does have an approximate standard normal distribution when μ_i is large. With standardized residuals it is easier to tell when a deviation $(y_i - \hat{\mu}_i)$ is “large.” Standardized residuals larger than about 2 or 3 in absolute value are worthy of attention, although some values of this size occur by chance alone when the number of observations is large.

Outline

- 1 3.1 Components of A Generalized Linear Model
- 2 3.2 Generalized Linear Models for Binary Data
- 3 3.3 Generalized Linear Models for Count Data
- 4 3.4 Statistical Inference and Model Checking
- 5 3.5 Fitting Generalized Linear Models
- 6 Homework 3

3.5.1 The Newton-Raphson Algorithm Fits GLMs

The Newton-Raphson Algorithm

The Newton-Raphson algorithm approximates the log-likelihood function in a neighborhood of the initial guess by a polynomial function that has the shape of a concave (mound-shaped) parabola. It has the same slope and curvature at the initial guess as does the log-likelihood function. It is simple to determine the location of the maximum of this approximating polynomial. That location comprises the second guess for the ML estimates. The algorithm then approximates the log-likelihood function in a neighborhood of the second guess by another concave parabolic function, and the third guess is the location of its maximum. The process is called iterative.

3.5.2 Wald, Likelihood-Ratio, and Score Inference Use the Likelihood Function

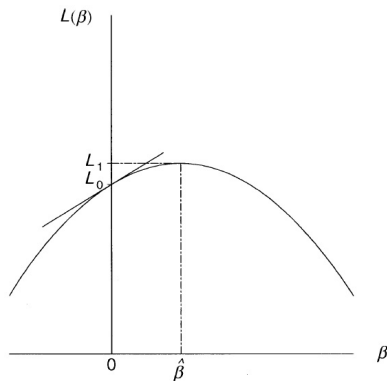


Figure 3.7. Information used in Wald, likelihood-ratio, and efficient score tests.

3.5.2 Wald, Likelihood-Ratio, and Score Inference Use the Likelihood Function

- Wald, $\hat{\beta}$, $(\hat{\beta}/SE)^2$. The SE of $\hat{\beta}$ depends on the curvature of the log-likelihood function at the point where it is maximized, with greater curvature giving smaller SE values.
- Score, log-likelihood function at the null value for β of 0. It uses the size of the derivative (slope) of the log-likelihood function, evaluated at the H_0 value of the parameter. The derivative at $\beta = 0$ tends to be larger in absolute value when $\hat{\beta}$ is further from that null value.

3.5.2 Wald, Likelihood-Ratio, and Score Inference Use the Likelihood Function

- LR, combines information about the log-likelihood function both at $\hat{\beta}$ and at $H_0 : \beta = 0$, using the chi-squared statistic $-2(L_0 - L_1)$. In a sense, this statistic uses the most information of the three types of test statistic. It is usually more reliable than the Wald statistic, especially when n is small to moderate.

3.5.3 Advantages of GLMs

The development of GLM theory in the mid-1970s unified important models for continuous and categorical response variables. Table 3.5 lists several popular GLMs for practical application. A nice feature of GLMs is that the model-fitting algorithm, Fisher scoring, is the same for any GLM. This holds regardless of the choice of distribution for Y or link function. Therefore, GLM software can fit a very wide variety of useful models.

3.5.3 Advantages of GLMs

Table 3.5 Types of Generalized Linear Models for Statistical Analysis

Random	Link	Systematic	Model	Chapter
Normal	Identity	Continuous	Regression	
Normal	Identity	Categorical	Analysis of variance	
Normal	Identity	Mixed	Analysis of covariance	
Binomial	Logit	Mixed	Logistic regression	4-5,8-10
Multinomial	Logits	Mixed	Multinomial response	6,8-10
Poisson	Log	Mixed	Loglinear	7

Quiz

Refer to Table 2.7 on x = mother's alcohol consumption and Y = whether a baby has sex organ malformation. With scores (0, 0.5, 1.5, 4.0, 7.0) for alcohol consumption, ML fitting of the linear probability model has the output:

Parameter	Estimate	Standard error	Likelihood 95%	Ratio CI
Intercept	0.00255	0.0003	0.0019	0.0032
Alcohol	0.00109	0.0007	-0.0001	0.0027

- State the prediction equation, and interpret the intercept and slope.
- Use the model fit to estimate: (i) probabilities of malformation for alcohol levels 0 and 7.0; (ii) relative risk comparing those two levels.

Outline

- 1 3.1 Components of A Generalized Linear Model
- 2 3.2 Generalized Linear Models for Binary Data
- 3 3.3 Generalized Linear Models for Count Data
- 4 3.4 Statistical Inference and Model Checking
- 5 3.5 Fitting Generalized Linear Models
- 6 Homework 3

Homework 3

1. Analysis the GDS5037 data. Suppose the samples are randomly chosen. Let Y denote patient's status, $Y = 1$ for severe asthma (SA), and $Y = 0$ for control.
 - (1) Use the identifier VPS39 as independent variable to built a logistic regression. What do you find? Explain your result.
 - (2) Use the identifier VPS39 as independent variable to built a probit model. What do you find? Explain your result.
2. Problems in textbook 3.3, 3.9, 3.10, 3.11.

Thank you!