Introduction to Logistic Regression Part I

Dr. Kiah Wah Ong

ILLINOIS TECH

Introduction

In MATH 764 and MATH 765:

(a) Our response variable y has always been a continuous quantitative variable.

(b) Our predictor variables, have been both

  (i) quantitative

  (ii) qualitative (recall the use of indicator variables)

However, the response variable can also be qualitative. In this module, we present methods for dealing with this situation.

Regression Models with Binary Response Variable

Example

Insurance possession study

- X_1 : age of head of household (quantitative)
- X_2 : amount of liquid assets (quantitative)
- X_3 : head of household's occupation (qualitative)
- Y : Household has insurance coverage or does not have insurance coverage.

The response variable Y can be codded as $Y = 0$ for not having an insurance coverage while $Y = 1$ as having insurance coverage.

Regression Models with Binary Response Variable

Example

Longitudinal study of coronary heart disease

- X_1 : age (quantitative)
- X_2 : gender (qualitative)
- X_3 : cholesterol level (quantitative)
- X_4 : BMI (quantitative)
- X_5 : blood pressure (quantitative)
- X_6 : smoking history (quantitative)
- Y : a person developed or did not develop heart disease during the study

The response variable Y can be codded as $Y = 0$ for not develop heart disease while $Y = 1$ for developing heart disease during the study.

New Technique is Needed

Constraints on Response Function

Let us consider a simple regression model with one predictor X.

Instead of predicting the two values of Y (either $Y = 0$ or $Y = 1$), let's model the probabilities that the response takes one of these two values. That is

Let π denote the probability that $Y = 1$ when $X = x$ and write

$$\pi = \Pr(Y = 1|X = x) = \beta_0 + \beta_1 x$$

This, however is problematic as π must lie between 0 and 1, while the linear function on the right is unbounded.

New Technique is Needed

Nonconstant Error Variance

Notice that Y is a Bernoulli random variable with the variance of Y given by $\pi(1-\pi)$. But since

$$\pi = \beta_0 + \beta_1 x,$$

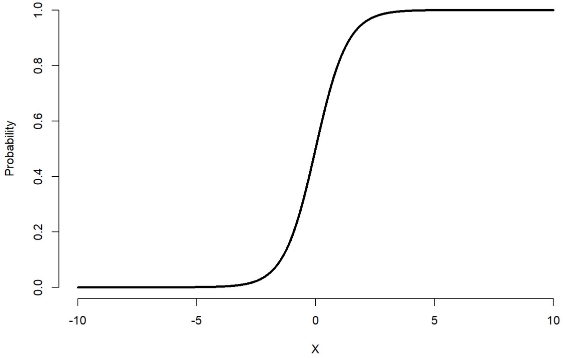
the variance of Y in this case is a function of x, hence the assumption of constant variance (homoscedasticity) does not hold.

Even though we can remediate the problem of unequal error variance (using weighted least squares), the difficulties created by unbounded linear function $\beta_0 + \beta_1 x$ is the most serious.

A new method for this problem is needed.

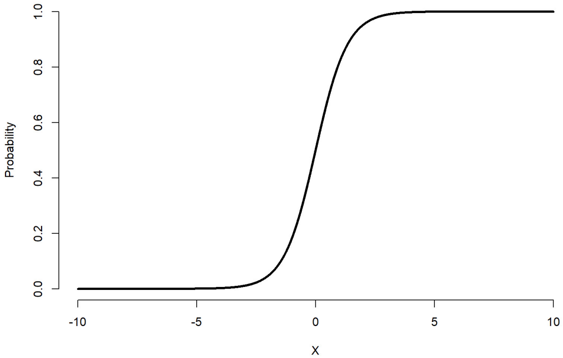
Dealing with Constraints on Response Function

We need a function for modeling the probability of the binary responses. The relationship between the probability π and X can be represented by a logistic response function shown below:



The function has a characteristic sigmoidal or S-shaped, and approach 0 and 1 asymptotically.

The Logit Model



The shape of the S-curve above can be produced if the model of the probability is given as:

$$\pi = \Pr(Y = 1|X = x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}}$$

The Logit Model

The logistic model we've seen above can be generalized easily to model with several predictive variables.

In the case of multiple predictive variables, the probability π is then modeled as

$$\begin{aligned}\pi &= \Pr(Y = 1 | X_1 = x_1, \cdots, X_p = x_p) \\ &= \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}\end{aligned}$$

The equation above is called the logistic regression function.

The Logit Model

Notice that the logistic regression function

$$\begin{aligned}\pi &= \Pr(Y = 1 | X_1 = x_1, \cdots, X_p = x_p) \\ &= \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}\end{aligned}$$

is nonlinear in the parameters $\beta_0, \beta_1, \cdots, \beta_p$.

However, it can be linearized by the logit transformation, as we will demonstrate next.

The Logit Model

Since

$$\begin{aligned}\pi &= \Pr(Y = 1|X_1 = x_1, \cdots, X_p = x_p)\\ &= \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}\end{aligned}$$

we see that

$$1-\pi = \Pr(Y = 0|X_1 = x_1, \cdots, X_p = x_p) = \frac{1}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}}$$

and

$$\frac{\pi}{1-\pi} = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}.$$

Taking the natural logarithm leads to

$$g(x_1, \cdots, x_p) = \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

The Logit Model

The ratio $\dfrac{\pi}{1-\pi}$ in the expression

$$\frac{\pi}{1-\pi} = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}.$$

is called the odds for the event.

Thus, the logistic regression is said to model the log-odds (logit) with a linear function of the predictors.

$$g(x_1,\cdots,x_p) = \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

Note: $\operatorname{logit}(p) = \ln\left(\dfrac{p}{1-p}\right)$

Odds

The concept of odds:

With $\pi(\mathbf{X}) = \pi = \Pr(Y = 1|\mathbf{X})$, the ratio

$$\begin{aligned}\frac{\pi}{1-\pi} &= \frac{\Pr(\mathrm{Y} = 1|\mathbf{X})}{1 - \Pr(Y = 1|\mathbf{X})} \\ &= \frac{\text{The number of "successful" event}}{\text{The number of "failure" event}}\end{aligned}$$

is called the odds of the "successful" event (the event corresponding to $Y = 1$).

Odds

Example

A box contains 5 white, 2 yellow and 6 red balls. What is the odds of drawing a white ball from the box?

In this case, a successful event corresponds to the event of drawing a white ball. Hence we have

$$\frac{\mathrm{Pr(Drawing\ a\ White\ ball)}}{1-\mathrm{Pr(Drawing\ a\ White\ ball)}} = \frac{5/13}{1-5/13} = \frac{5}{8} = 0.625$$

The odds are said to be 5:8. That is, on average the successful event will occur 5 times for every 8 times it does not.

Odds

Example

Referring to the previous example, that is, we have a box that contains 5 white, 2 yellow and 6 red balls.

If we ask, what is the odds of drawing a ball that is not white from the box, then the odds will be

$$\frac{\mathrm{Pr}(\text{Drawing a Nonwhite ball})}{1 - \mathrm{Pr}(\text{Drawing a Nonwhite ball})} = \frac{8/13}{1 - 8/13} = \frac{8}{5} = 1.6$$

From here, we see that odds, unlike probability can take values greater than one.

End of Part I

Next, we will see how to obtain the estimates $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p$.