Discrete Response Model
Lecture 3

datascience@berkeley

Generalized Linear Model

Generalized Linear Model

Logistic regression models fall within a family of models called *generalized linear models*. Each generalized linear model has three different components:

1. **Random:** This specifies the distribution for Y. For the logistic regression model, Y has a Bernoulli distribution.
2. **Systematic:** This specifies a linear combination of the β parameters with the explanatory variables, and it is often referred to as the linear predictor. For the logistic regression model, we have $\beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$.
3. **Link:** This specifies how the expected value of the random component E(Y) is linked to the systematic component. In logistic regression model, logit is the link function.

Note that "linear" in generalized linear models comes from the β parameters simply being coefficients for the explanatory variables in the model. Nonlinear models involve more complex functional forms such as x^{β}.

Probit Regression Model

While the logit-link function is the most prevalently used for binary regression, there are two other functions that are common:

Inverse CDF of a standard normal distribution: This produces what is known as a probit regression model.

Suppose $\Phi(\cdot)$ denotes the CDF of a standard normal distribution. The model is written as

$$\pi = \Phi(\beta_0 + \beta_1 x_1 + \cdots \beta_p x_p)$$

or equivalently as

$$\boxed{\Phi^{-1}(\pi) = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p}$$

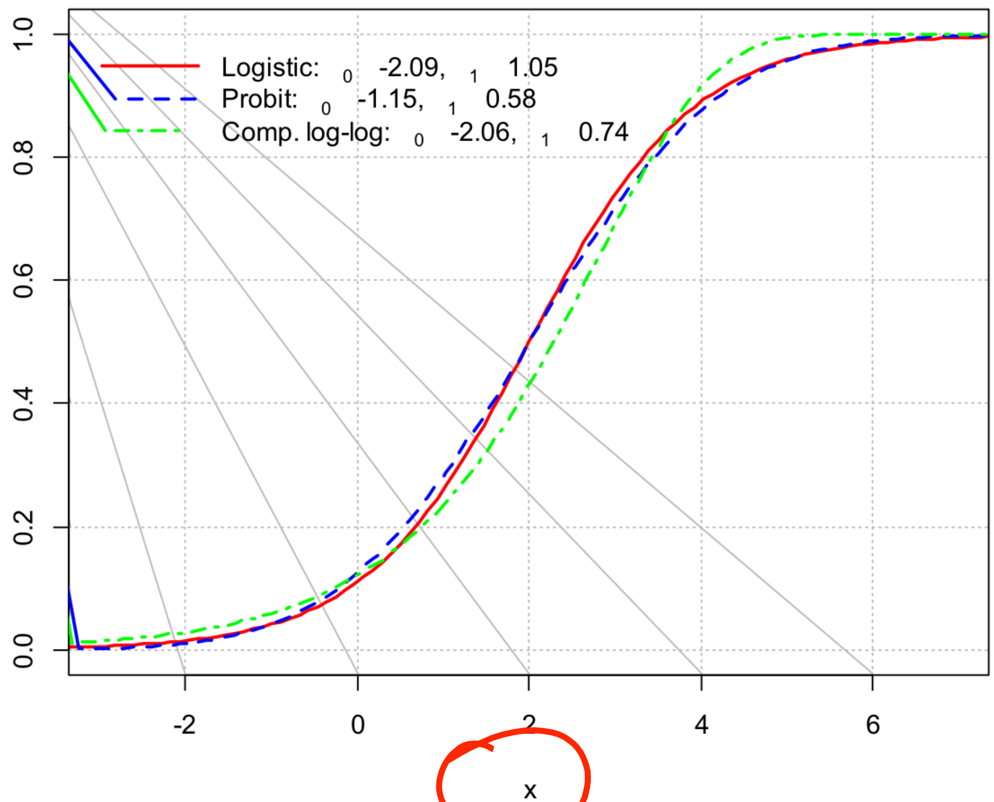
A very common way to express the model is

$$\text{probit}(\pi) = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$$

where probit is used to denote the inverse CDF transformation in a similar manner as the logit transformation is for logistic regression.

Comparison of Three Types of Models

Suppose the linear predictor has only one explanatory variable of the form $\beta_0 + \beta_1 x$. By choosing values of β_0 and β_1 so that the mean is 2 and the variance is 3 for the corresponding CDFs, we obtain the plots of the models displayed below (see the corresponding program for code):



Comparison of Three Types of Models

Estimation: Probit and complementary log-log models are estimated in the same way as the logistic regression model. The difference now is that π is represented in the log-likelihood function by the corresponding probit or complementary log-log model specification.

Inference: Once the parameter estimates are found, the same inference procedures as used for the logistic regression model are available for the probit and complementary log-log models.

Odds ratios: Odds ratios are not as easy to interpret with probit and complementary log-log regression models as they were for logistic regression models. The same simplification does not hold true for probit and complementary log-log models. This is one of the main reasons why logistic regression models are the most used binary regression models.

Comparison of Three Types of Models

Example: Odds ratios used with probit models

Consider the model probit(π) = $\beta_0 + \beta_1 x$. The odds of a success are

$$\text{Odds}_x = \Phi(\beta_0 + \beta_1 x) / \left[1 - \Phi(\beta_0 + \beta_1 x)\right]$$

at a particular value of x. The odds of a success with a c-unit increase in x are

$$\text{Odds}_{x+c} = \Phi(\beta_0 + \beta_1 x + \beta_1 c) / \left[1 - \Phi(\beta_0 + \beta_1 x + \beta_1 c)\right]$$

When the odds ratio is formed from these two odds, x will remain in the final expression! Therefore, the odds ratio for a probit regression model depends on the value of the explanatory variable.

Berkeley SCHOOL OF INFORMATION