Noname manuscript No.
(will be inserted by the editor)

Logit and probit models

Kelvin Yang Li

Started: 27 October 2018/ Finished: 27 October 2018

A *sigmoid function* is a function having a S-shaped curve. It is monotonically increasing function that takes inputs between $(-\infty, \infty)$ and outputs between $(0, 1$ or $(-1, 1)$. Its mathematical form is

$$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}.$$

Examples of sigmoid functions are the logistic function, hyperbolic tangent function, CDFs, etc.

1 Logit model

The *logistic distribution* is a continuous probability distribution that has bell shaped PDF curves. Its PDF is parameterized by mean μ and a scale parameter $s = q\sigma$ that is proportional to the standard deviation by $q = \sqrt{3}/\pi$. The PDF of the logistic distribution is expressed as

$$f(x;\mu;s) = \frac{e^{-\frac{x-\mu}{s}}}{s(1+e^{-\frac{x-\mu}{s}})^2} = \frac{1}{s(e^{\frac{x-\mu}{2s}}+e^{-\frac{x-\mu}{2s}})^2}.$$

The CDF of the logistic distribution is also known as the *logistic function* and is expressed as

$$F(x;\mu;s) = \frac{1}{1+e^{-\frac{x-\mu}{s}}},$$

Kelvin Yang Li
Faculty of IT, Monash University, Victoria, Australia
E-mail: yang.kelvinli@monash.edu

which is a function of the family of the logistic functions. The standard logistic function is when $\mu = 0$ and $s = 1$, so it has the same functional form as the sigmoid function. The logistic function often mentioned when introducing logit model refers to the standard logistic function (or the sigmoid function).

In a logit model, each predictor x_i has an associated parameter β_i, so the *log-odds* $t = \beta_0 x_0 + \cdots + \beta_m x_m = \boldsymbol{\beta} \cdot \mathbf{x} \in (-\infty, \infty)$ for $i \in [0, m]$, where β_0 is known as the intercept and setting $x_0 = 1$ is a trick for both programming convenience and expressing t in a compact form. Since probability=odds/(1+odds), the logistic function converts the log-odds into probabilities by

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-\boldsymbol{\beta} \cdot \mathbf{x}}}.$$

The observed data for logit models are xs and and a binary y. We can assume there exists a latent variable $y^* = \boldsymbol{\beta} \cdot \mathbf{x} + \epsilon$, where ϵ is an unobserved error term but assumed to follow a standard logistic distribution. Hence, y can be interpreted as an indicator of whether or not y^* is positive. That is, if $\boldsymbol{\beta} \cdot \mathbf{x} + \epsilon > 0$ then $y = 1$, else $y = 0$. Hence, we have

$$\begin{aligned} P(y = 1 \mid x) &= p(\epsilon > -\boldsymbol{\beta} \cdot \mathbf{x}) \\ &= p(\epsilon < \boldsymbol{\beta} \cdot \mathbf{x}) \\ &= \frac{1}{1 + e^{-\boldsymbol{\beta} \cdot \mathbf{x}}}. \end{aligned}$$

The logit model is a case of the family called *generalized linear model.* It uses the standard logistic function as the link function.

2 Probit model

The probit model is also in the family of generalized linear model, but uses a different link function from the logit model. In particular, the probit model uses the CDF $\Phi()$ of the standard normal distribution. The functional form of the probit model $P(y = 1 \mid x) = \Phi(\mathbf{beta} \cdot \mathbf{x})$ can also be derived using a latent variable y^* and an unobserved error $\epsilon \sim N(0, 1)$ that assumed to follow a standard normal distribution.