Link Function Alternatives For Logistic Regression

M.Sinan İyisoy

Necmettin Erbakan University

Meram Faculty of Medicine

Generalized Linear Models

- Nelder and Wedderburn introduced Generalized Linear Models in 1972.
- In this framework, such a model is specified by

i) independent observations y_i distributed according to an exponential family distribution

ii) a set of explanatory variables x, for each observation, forming systematic component

iii) The link function $g(\mu) = Y$ relating mean of an observation to systematic component.

Logistic Regression

As a special case of Generalized Linear Model, logistic regression model has components

1) Random component

Response variable y and its probability distribution where y is a dichotomous variable with Bernoulli distribution.

2) Linear Predictor (Systematic component)

A linear combination of explanatory variables of different types that is linear in the parameters

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + .. + \beta_p x_{ip}$$

Logistic Regression

3) Link Function

A smooth and invertible function g(.) that transforms π_i to the linear predictor.

$$g(\pi_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + .. + \beta_p x_{ip}$$

So $\pi_i = g^{-1}\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + .. + \beta_p x_{ip}\right)$

g^{-1} should map the real line onto (0,1) because probabilities π_i obtained must be between 0 and 1.

Cumulative Distribution Functions

- Cumulative distribution functions of most continous distributions (generally denoted by F) has the following appealing properties
- Mapping real line onto (0,1)
- Non-decreasing (S-shaped)
- Invertibility (Quantile functions).

Quantile Functions

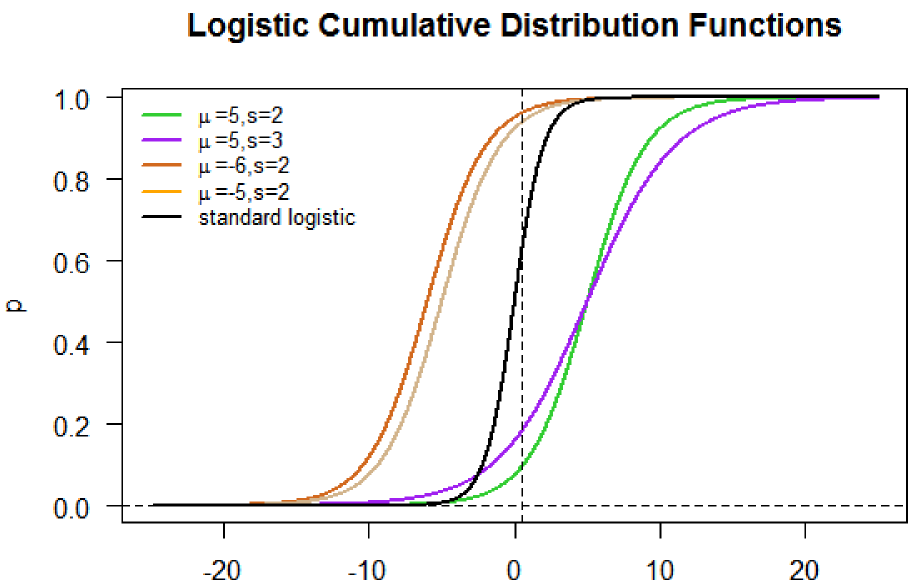
- Inverse of a CDF is called **Quantile** function. A Quantile function is a candidate for link function in a logistic regression model.
- For logistic distribution, its CDF is

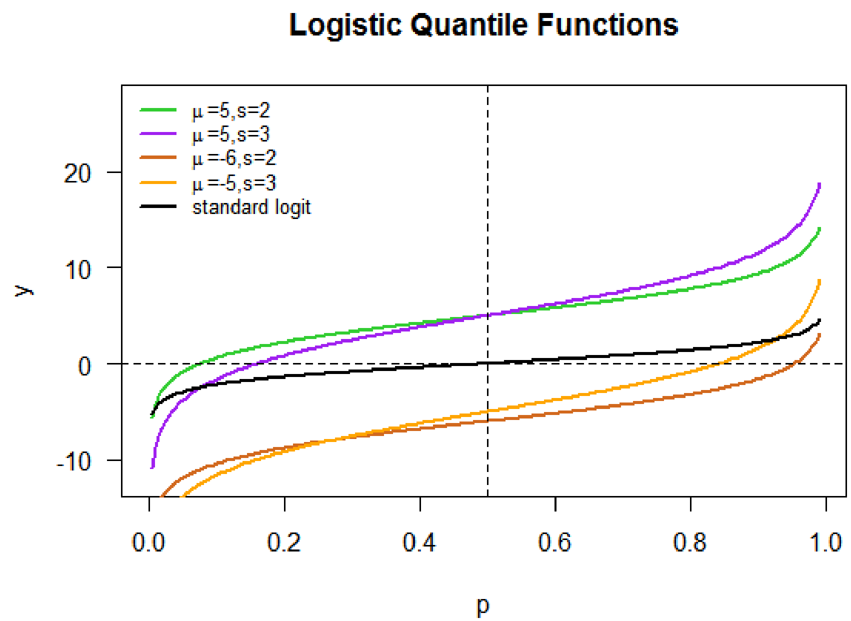
$F(x; \mu, s) = \frac{1}{1+e^{-\frac{x-\mu}{s}}}$ for parameters μ, s and its quantile function is a generalisation of logit

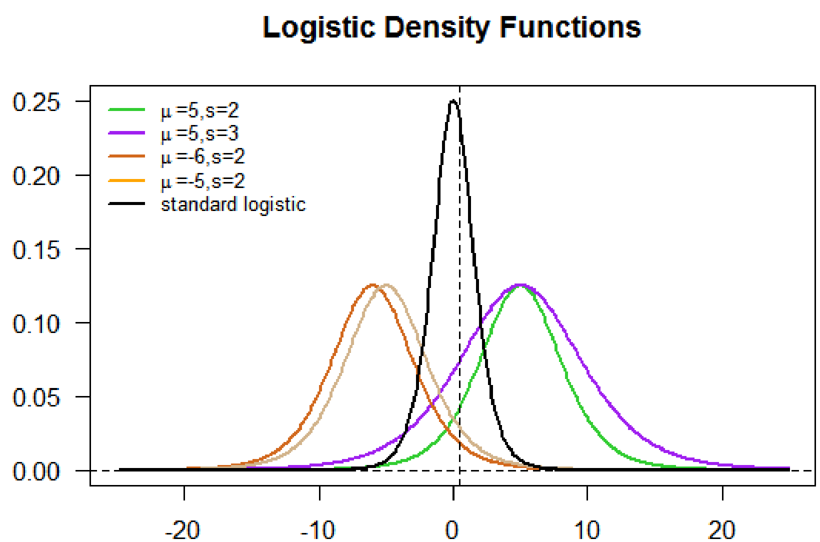
$$Q(p; \mu, s) = \mu + s\ log(\frac{p}{1-p})$$

Taking $\mu = 0\ and\ s = 1$, we get the standard logit function. Notice that standard logit Q takes a value p in (0,1) and returns a real number.

Logistic Distribution







Canonical Link (Berkson 1944)

The canonical link function used in logistic regression model is the standard logit function

$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) \ .$$

Since g is invertible, we have

$$g^{-1}(\eta_i) = logistic(\eta_i) = \frac{1}{1 + e^{-\eta_i}}$$

How to find other g's?

- One of the important aspects when choosing a link function is its symmetry about 0.5,

  ie. $g(x)=-g(1-x)$.
- As seen from previous graphs, if probability distribution is symmetric about 0 then its cumulative distribution and quantile functions are symmetric about 0.5.
- Symmetric link functions can be obtained from symmetric density functions like standard normal, logistic, Student-t and Cauchy.

Asymmetric Links

- Similarly, asymmetric links can be obtained from cdfs of asymmetric distributions.
- There are other ways to obtain asymmetric link functions:
- i) modifying linear predictor η_i, making it nonlinear
- ii) considering F in a more general parametric class of probability distributions.
- A well-known example to an asymmetric link function is complementary log-log due to Fisher (1922).

Latent Variable Model

- Each link function corresponds to a latent variable model

$y_i^* = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + .. + \beta_p x_{ip} + \epsilon_{ip}$

or in short

$Y_i^* = X_i \beta + \epsilon_i$ (assuming there is an underlying continous latent variable Y^*)

where ϵ_i is iid and $Y_i = I(Y_i^* > 0)$. Namely

$$\begin{cases} Y_i = 1 \ \ if \ Y_i^* > 0 \\ Y_i = 0 \ \ if \ Y_i^* \leq 0 \end{cases}$$

Maximum likelihood estimation of the latent variable model implies a link function chosen as the quantile function F of ϵ_i's.

Latent Variable Model

Since

$$p_i = P[y_i = 1|X_i] = \Pr(X_i\beta + \epsilon_i > 0)$$

$= \Pr(\ \epsilon_i > -X_i\beta) = 1 - \Pr(\ \ \epsilon_i < -X_i\beta)$

$= 1 - F(-X_i'\beta)$

we obtain $p_i = 1 - F(-X_i'\beta)$. So

$F^{-1}(p_i) = F^{-1}\big(1 - F(-X_i'\beta)\big) =$

$-F^{-1}(F(-X_i'\beta)\ ) = X_i'\beta$. Hence $g = F^{-1}$

constitutes a link function.

Some Other Link Functions

- Probit
- Clog-log
- Log-log
- Cauchit
- Robit
- Scobit
- Pregibit

Probit (Bliss 1934)

- The quantile function of standard normal distribution is called probit and denoted by $\Phi^{-1}$where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$.
- Hence $\Phi^{-1}(\pi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + .. + \beta_p x_{ip}$

Here ϵ_i 's are supposed to have a standard normal density function so the link is chosen as Φ^{-1}.

CLog-Log (Fisher 1922)

- Complementary Log-Log function is defined to be $g(\pi_i) = \log(-\log(1-\pi_i))$ which is the inverse cdf (quantile) of Gumbel distribution.
- $g^{-1}(\eta_i) = 1 - e^{(-e^{\eta_i})}$
- In this case ϵ_i ’s are supposed to have a Gumbel density function.

Log-Log

- Log-Log function is defined to be
- $g(\pi_i) = -\log(-\log(\pi_i))$ which is the inverse cdf (quantile) of Log-Weibull distribution.
- $g^{-1}(\eta_i) = e^{(-e^{\eta_i})}$
- Log-log and Clog-log models are mirror images of each other with the same kurtosis and opposite skewness.

Complementary Log (Piegorsch 1992)

- Complementary Log function is defined to be $g(\pi_i) = -\log(1 - \pi_i)$ which is the quantile of standard exponential distribution.
- $g^{-1}(\eta) = 1 - e^{-\eta}$ defined for $\eta > 0$.
- Used for truncated data.

Cauchit

- Cauchit is defined to be

$$g(\pi_i) = \tan(\pi\left(\pi_i - \frac{1}{2}\right))$$

which is the quantile of Cauchy distribution.

- $g^{-1}(\eta) = \frac{1}{\pi}\arctan(\eta) + \frac{1}{2}$
- Since Cauchy distribution is a member of Student t family distributions where degree of freedom $v = 1$, Cauchit is a member of Gosset Link family.

Robit (Liu 2004)

- The quantile function of Student t distribution with v degrees of freedom is called robit and denoted generically by F_v^{-1} where F_v is the cumulative distribution function of Student t distribution with $df = v$.
- Robit link with 7 df approximates logistic link and Robit link with $df \to \infty$ is probit.
- Robit regression is allegedly robust.

Scobit (Skewed Logit, Nagler 1994)

- Scobit is defined to be

$$g(\pi_i) = log((1-\pi_i)^{-\frac{1}{\lambda}} - 1)$$

which is the quantile of Burr type II distribution for some parameter λ where

$$g^{-1}(\eta) = 1 - \frac{1}{(1+e^{\eta})^{\alpha}} \quad .$$

Power Logit

- Power Logit is defined to be

$$g(\pi_i) = -log(\pi_i^{-\frac{1}{\lambda}} - 1)$$

which is the quantile of exponential Burr type XII distribution for some parameter λ where

$$g^{-1}(\eta) = \frac{1}{(1+e^{-\eta})^{\alpha}} \quad .$$

Burr Distributions (Burr 1942)

$$F(x) = x, \qquad (0,\ 1), \tag{9}$$

$$F(x) = (e^{-x} + 1)^{-r}, \tag{10}$$

$$F(x) = (x^{-k} + 1)^{-r}, \qquad (0,\ \infty), \tag{11}$$

$$F(x) = \left[\left(\frac{c - x}{x}\right)^{1/c} + 1\right]^{-r}, \qquad (0,\ c), \tag{12}$$

$$F(x) = (ke^{-\tan x} + 1)^{-r}, \qquad \left(-\frac{\pi}{2},\ \frac{\pi}{2}\right), \tag{13}$$

$$F(x) = (ke^{-c \sinh x} + 1)^{-r}, \tag{14}$$

$$F(x) = 2^{-r}(1 + \tanh x)^{r}, \tag{15}$$

$$F(x) = \left(\frac{2}{\pi} \arctan e^{x}\right)^{r}, \tag{16}$$

$$F(x) = 1 - \frac{2}{k[(1 + e^{x})^{r} - 1] + 2}, \tag{17}$$

$$F(x) = (1 - e^{-x^{2}})^{r}, \qquad (0,\ \infty), \tag{18}$$

$$F(x) = \left(x - \frac{1}{2\pi} \sin 2\pi x\right)^{r}, \qquad (0,\ 1), \tag{19}$$

$$F(x) = 1 - (1 + x^{c})^{-k}, \qquad (0,\ \infty), \tag{20}$$

Gosset Link Family (Koenker&Yoon 2009)

- Gosset link family includes link functions being quantiles of Student t distributions of different df.
- Probit and cauchit links are naturally nested within Student t Family.
- There is an extensive literature on the use of Student models for continous response data where it has robustness advantages.

Gosset Link Family

- When $v > 6$ it is difficult to distinguish Gosset models from probit.
- For $v = 7\ or\ v = 8$, the logit model is well approximated by Gosset model.
- When $v < 0.2$ evaluation of likelihood becomes problematic.

Pregibit (Vijverberg 2012)

- The quantile function

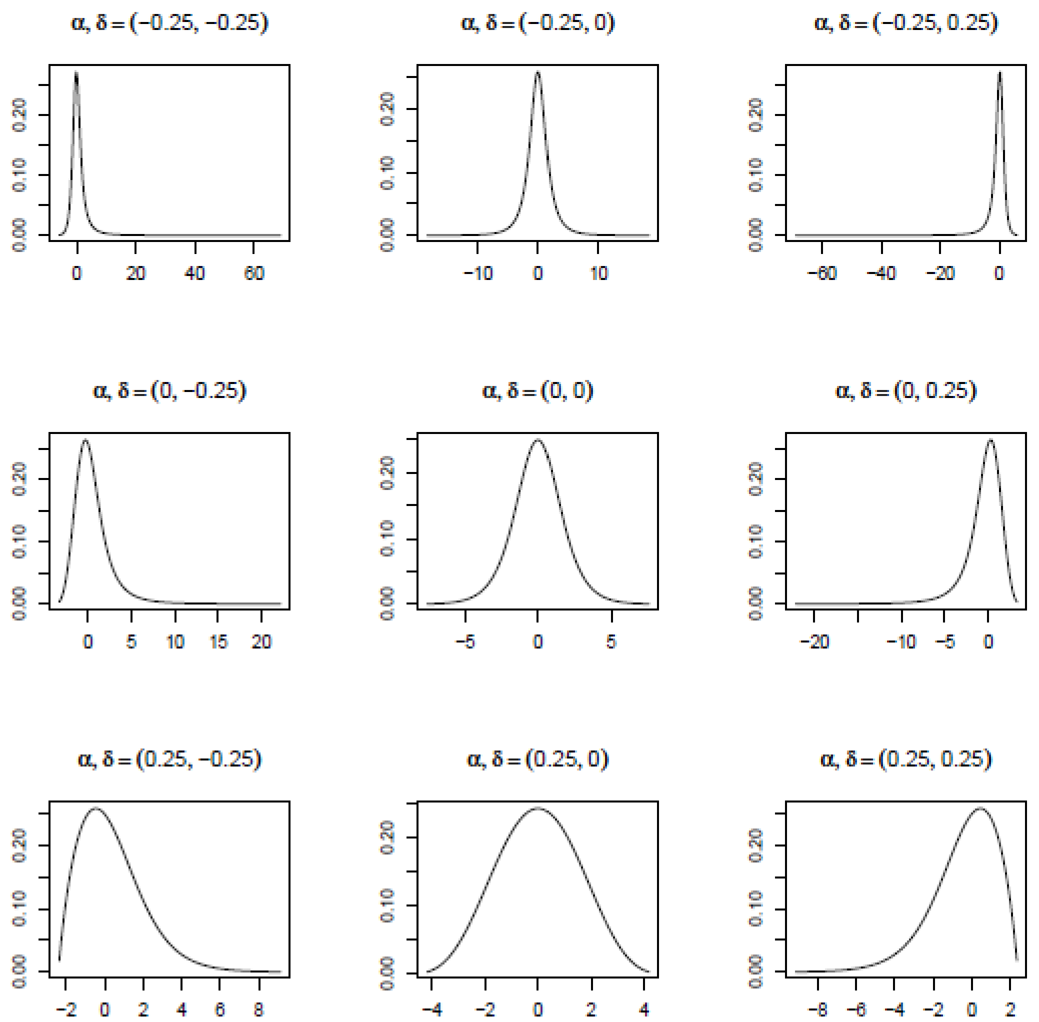
$$g(\pi_i; \alpha, \delta) = \frac{\pi_i^{\alpha-\delta} - 1}{\alpha - \delta} - \frac{(1-\pi_i)^{\alpha-\delta} - 1}{\alpha + \delta}$$

of Generalized Tukey Lambda distribution $GTL(\alpha, \delta)$ with parameters α, δ is called pregibit.

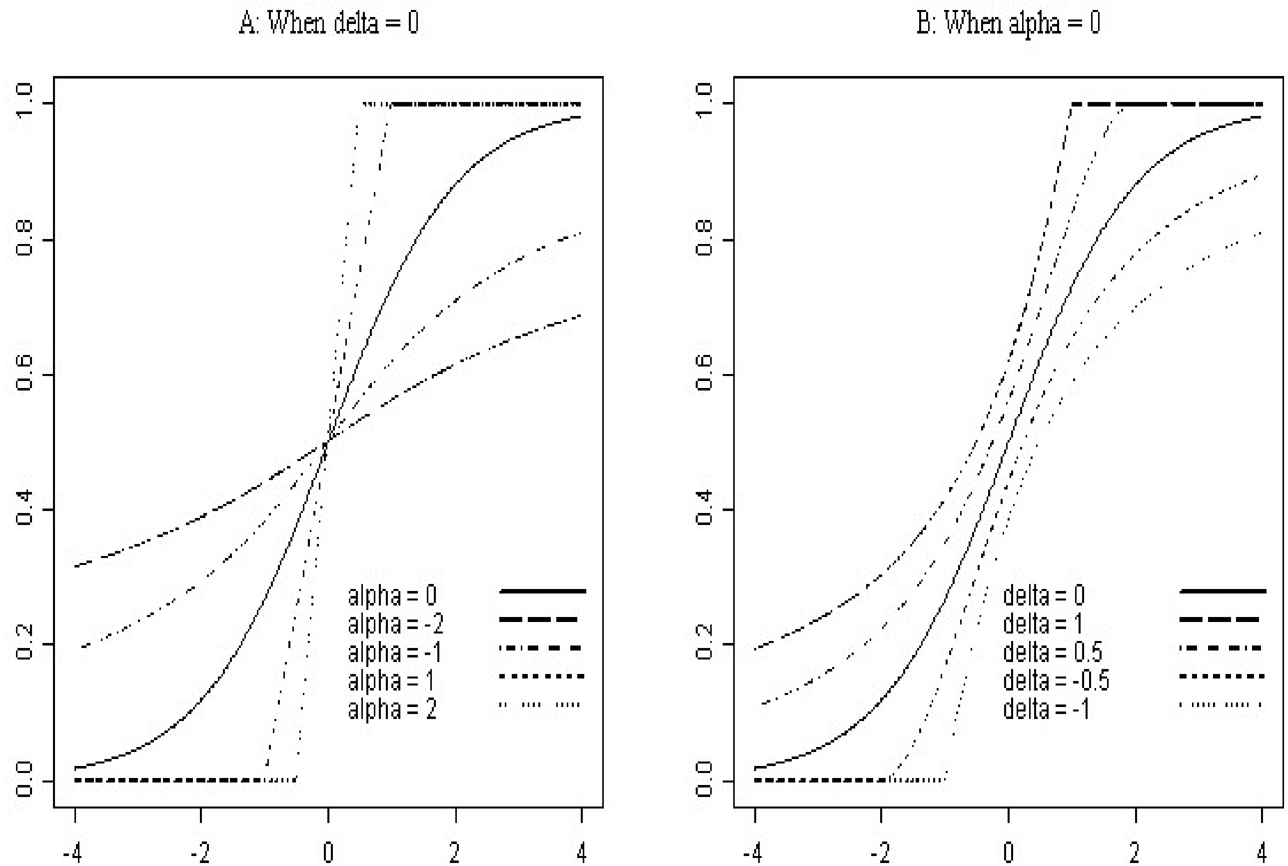
This link function generates diverse family of GTL density functions.

The pregibit model is simply a binary choice model with errors following GTL distributions.

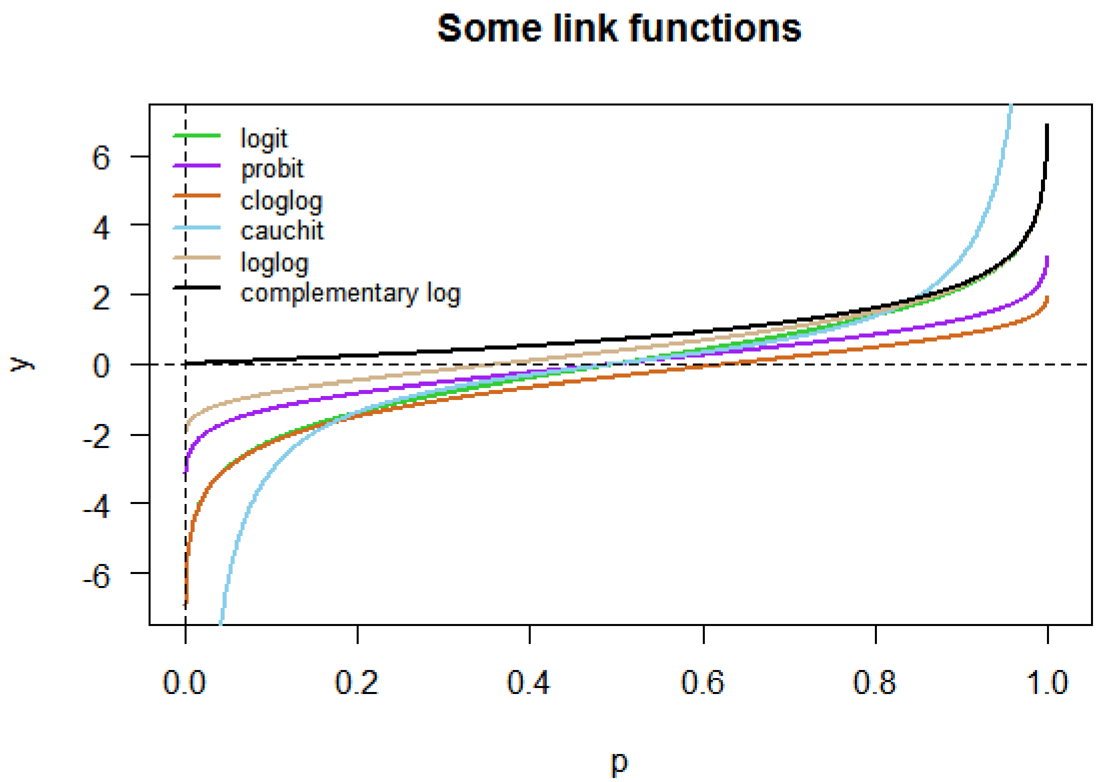
Some GTL Distributions



Some Pregibit CDFs



Graphs of Some Links



Thank you