ECON 452* -- NOTE 14

Maximum Likelihood Estimation of Binary Dependent Variables Models: Probit and Logit

This note demonstrates how to formulate binary dependent variables models for maximum likelihood estimation, and how to estimate by maximum likelihood the two most common formulations of such models, namely probit and logit models.

1. General Formulation of Binary Dependent Variables Models

A conventional formulation of binary dependent variables models relates the **observed binary outcome variable Y_i** to an ***unobserved* (or *latent*) dependent variable** Y_i^*.

- The ***unobserved* (or *latent*) dependent variable** Y_i^* is assumed to be generated by a classical linear regression model of the form

$$Y_i^* = x_i^T\beta + u_i \qquad (1)$$

where:

Y_i^* = a continuous real-valued index variable for observation i that is *unobservable*, or *latent*;

$x_i^T = (1\ X_{i1}\ X_{i2}\ \cdots\ X_{ik})$, a 1×K row vector of regressor values for observation i;

$\beta = (\beta_0\ \beta_1\ \beta_2\ \cdots\ \beta_k)^T$, a K×1 column vector of regression coefficients;

u_i = an iid random error term for observation i.

- The **random error terms u_i** are assumed to have zero conditional means and constant conditional variances for any set of regressor values x_i^T:

$$E\left(u_i \middle| x_i^T\right) = 0 \qquad \forall \ i \tag{2.1}$$

$$Var\left(u_i \middle| x_i^T\right) = E\left(u_i^2 \middle| x_i^T\right) = \sigma^2 \qquad \forall \ i \tag{2.2}$$

In addition, the conditional distribution of the u_i is assumed to be ***symmetric*** **around their zero conditional mean**.

Symmetry around mean zero means that

$$\Pr(u_i \le -a) = \Pr(u_i > a)$$

Since by definition $\Pr(u_i > a) = 1 - \Pr(u_i \le a)$, symmetry means that

$$\Pr(u_i \le -a) = 1 - \Pr(u_i \le a) \quad \textit{or} \quad \Pr(u_i \le a) = 1 - \Pr(u_i \le -a). \tag{2.3}$$

- The ***observable outcomes*** **of the binary choice problem** are represented by a **binary indicator variable Y_i** that is related to the unobserved dependent variable Y_i^* as follows:

$$Y_i = 1 \text{ if } Y_i^* > 0 \tag{3.1}$$

$$Y_i = 0 \text{ if } Y_i^* \leq 0 \tag{3.2}$$

The **random indicator variable Y_i** represents the observed realizations of a binomial process with the following probabilities:

$$\Pr(Y_i = 1) = \Pr(Y_i^* > 0) = \Pr(x_i^T\beta + u_i > 0) \tag{5.1}$$

$$\Pr(Y_i = 0) = \Pr(Y_i^* \leq 0) = \Pr(x_i^T\beta + u_i \leq 0) \tag{5.2}$$

What is required to estimate the coefficient vector β are analytical representations of the binomial probabilities (5.1) and (5.2).

- **Interpretation of the regression function**

  Under the zero conditional mean error assumption (2.1), equation (1) implies that

$$E\left(Y_i^* \middle| x_i^T\right) = E\left(x_i^T\beta \middle| x_i^T\right) + E\left(u_i \middle| x_i^T\right) = x_i^T\beta. \tag{4}$$

♦ The ***regression function*** $x_i^T\beta$ is thus the conditional mean value of the latent random variable Y_i^* for given values of the regressors.

♦ The ***slope coefficients*** $\boldsymbol{\beta_j}$ (j = 1, …, k) are the partial derivatives of the regression function (4) with respect to the individual regressors:

$$\frac{\partial E\left(Y_i^* \middle| x_i^T\right)}{\partial X_{ij}} = \frac{\partial x_i^T\beta}{\partial X_{ij}} = \frac{\partial(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_j X_{ij} + \cdots + \beta_k X_{ik})}{\partial X_{ij}} = \beta_j.$$

2. Analytical Representation of Binomial Probabilities

The binomial probabilities

$$\Pr(Y_i = 1) = \Pr(Y_i^* > 0) = \Pr(x_i^T\beta + u_i > 0) \qquad (5.1)$$

$$\Pr(Y_i = 0) = \Pr(Y_i^* \le 0) = \Pr(x_i^T\beta + u_i \le 0) \qquad (5.2)$$

are represented analytically in terms of the ***cumulative distribution function***, or ***c.d.f.***, for the **random error term u_i in regression equation (1)**:

$$Y_i^* = x_i^T\beta + u_i \qquad (1)$$

- The ***cumulative distribution function* (*c.d.f.*)** for the random variable u is denoted in general by G(u) and is defined as

$$G(a) = \Pr(u \le a) = \int_{-\infty}^{a} g(u)du = \int_{-\infty}^{a} g(u)\,du$$

where

$$G(-\infty) = \Pr(u \le -\infty) = \int_{-\infty}^{-\infty} g(u)du = 0$$

$$G(\infty) = \Pr(-\infty \le u \le \infty) = \int_{-\infty}^{\infty} g(u)du = 1$$

$$G(a) \le G(b) \qquad \text{for } a < b$$

- The probability that $\Pr(u > a) = \Pr(u \ge a)$ is given in terms of G(a) as

$$\Pr(u > a) = G(\infty) - G(a) = 1 - G(a)$$

- For a < b, the probability $\Pr(a \le u \le b)$ is given as:

$$\Pr(a \le u \le b) = G(b) - G(a).$$

- The **first derivative of the c.d.f.** equals the corresponding ***probability density function***, or ***p.d.f.***:

$$g(u) = \frac{dG(u)}{du} \qquad \text{or} \qquad g(a) = \left.\frac{dG(u)}{du}\right|_{u=a} = \frac{dG(a)}{da}$$

where g(a) is the value of $dG(u)/du$ evaluated at u = a.

- The ***probability density function* (*p.d.f.*)** for the random variable u is the function g(u) defined over all real values of u such that:

1. $g(u) \geq 0$

2. $\int_{-\infty}^{\infty} g(u)du = 1$

3. for any real values a and b where $-\infty < a < b < \infty$,

$$\Pr(a \leq u \leq b) = \int_a^b g(u)du$$

- **<u>Symmetry Property</u>:** In addition to the assumptions that the random variable u has zero mean and constant (finite) variance σ^2, it is assumed that the p.d.f. g(u) is ***symmetric* about its *zero* mean**.

- ***Symmetry* of g(u) around mean zero** means that

$$g(-a) = g(a) \qquad \textit{and} \qquad \Pr(u \le -a) = \Pr(u > a).$$

Since by definition

$$\Pr(u \le -a) = G(-a) \qquad \textit{and} \qquad \Pr(u > a) = 1 - \Pr(u \le a) = 1 - G(a),$$

***symmetry* of g(u)** implies that

$$G(-a) = 1 - G(a) \qquad \textit{or equivalently that} \qquad G(a) = 1 - G(-a).$$

- **Geometrically**, the **symmetry property** means that **the lower tail area probability that u ≤ –a *is equal to* the upper tail area probability that u > a**.

***lower* tail area Pr(u ≤ –a) = *upper* tail area Pr(u > a)**

- **<u>Representation of the Binomial Probabilities</u>**

♦ The binomial probability $\Pr(Y_i = 1) = \Pr(Y_i^* > 0) = \Pr(x_i^T\beta + u_i > 0)$ can be represented in terms of the c.d.f. for the random variable u as follows:

$$\begin{aligned} \Pr(Y_i = 1) &= \Pr(Y_i^* > 0) \\ &= \Pr(x_i^T\beta + u_i > 0) \\ &= \Pr(u_i > -x_i^T\beta) \\ &= 1 - \Pr(u_i \le -x_i^T\beta) \\ &= 1 - G(-x_i^T\beta) \\ &= G(x_i^T\beta) \qquad \text{by symmetry of } g(u) \end{aligned} \qquad \textbf{(6.1)}$$

♦ The binomial probability $\Pr(Y_i = 0) = \Pr(Y_i^* \le 0) = \Pr(x_i^T\beta + u_i \le 0)$ can be represented in terms of the c.d.f. for the random variable u as follows:

$$\begin{aligned} \Pr(Y_i = 0) &= \Pr(Y_i^* \le 0) \\ &= \Pr(x_i^T\beta + u_i \le 0) \\ &= \Pr(u_i \le -x_i^T\beta) \\ &= G(-x_i^T\beta) \\ &= 1 - G(x_i^T\beta) \qquad \text{by symmetry of } g(u) \end{aligned} \qquad \textbf{(6.2)}$$

♦ The ***probability density function*, or *p.d.f.*, for the binary dependent variable Y_i** can thus be written as:

$$g(Y_i) = \left[G(x_i^T\beta)\right]^{Y_i}\left[1 - G(x_i^T\beta)\right]^{1-Y_i} \qquad \text{for } Y_i = 0, 1. \qquad (7)$$

3. The Sample Likelihood and Log-Likelihood Functions

♦ The **sample *likelihood* function** for a sample of N independent observations $\{Y_i : i = 1, \ldots, N\}$ is:

$$
\begin{aligned}
L(Y_1, Y_2, \ldots, Y_N) &= \prod_{i=1}^{N} g(Y_i) \\
&= \prod_{i=1}^{N} \left[G(x_i^T\beta)\right]^{Y_i} \left[1 - G(x_i^T\beta)\right]^{1-Y_i} \qquad (8) \\
&= \prod_{Y_i=1} G(x_i^T\beta) \prod_{Y_i=0} \left(1 - G(x_i^T\beta)\right)
\end{aligned}
$$

♦ The **sample *log-likelihood* function** for a sample of N independent observations $\{Y_i : i = 1, \ldots, N\}$ is:

$$
\begin{aligned}
\ln L(Y_1, Y_2, \ldots, Y_N) &= \ln(L) \\
&= \sum_{i=1}^{N} \ln g(Y_i) \\
&= \sum_{i=1}^{N} \left\{ Y_i \ln G(x_i^T\beta) + (1 - Y_i) \ln\left[1 - G(x_i^T\beta)\right] \right\} \\
&= \sum_{i=1}^{N} Y_i \ln G(x_i^T\beta) + \sum_{i=1}^{N} (1 - Y_i) \ln\left[1 - G(x_i^T\beta)\right] \qquad (9) \\
&= \sum_{Y_i=1} \ln G(x_i^T\beta) + \sum_{Y_i=0} \ln\left[1 - G(x_i^T\beta)\right]
\end{aligned}
$$

4. Distributional Specifications of the Model

- To complete specification of the model, a specific probability distribution must be chosen for the random error terms u_i.

  The most commonly adopted distributions in econometric applications are the **standard normal** and the **standard logistic**.

  1. The **standard *normal* distribution** yields the ***probit* model**.
  2. The **standard *logistic* distribution** yields the ***logit* model**.

Probit Model

- The **standard normal distribution** has mean $\mu = 0$ and variance $\sigma^2 = 1$, and is *symmetric* around its zero mean.

  If the random variable x_i is normally distributed with mean μ and variance σ^2, then the standard normal variable $z_i = (x_i - \mu)/\sigma$ is normally distributed with mean 0 and variance 1. That is,

  if $x_i \sim N(\mu, \sigma^2)$, then $z_i \sim N(0,1)$ where $z_i = (x_i - \mu)/\sigma$.

- The ***standard normal p.d.f.*** is

$$\phi(z_i) = (2\pi)^{-1/2} \exp\left(-\frac{z_i^2}{2}\right).$$

- The ***standard normal c.d.f.*** is

$$\Phi(Z_i) = \Pr(z \le Z_i) = \int_{-\infty}^{Z_i} \phi(z)dz = \int_{-\infty}^{Z_i} (2\pi)^{-1/2} \exp\left(-\frac{z^2}{2}\right)dz.$$

- Choice of the **standard normal** for the distribution of the random error terms u_i leads to the ***probit* model**.

Logit Model

- The **standard logistic distribution** has mean $\mu = 0$ and variance $\sigma^2 = \pi^2/3$, and is *symmetric* around its zero mean.

- The ***standard logistic p.d.f.*** is

$$f(x_i) = \frac{\exp(x_i)}{\left(1 + \exp(x_i)\right)^2} = \frac{\exp(-x_i)}{\left(1 + \exp(-x_i)\right)^2}.$$

- The ***standard logistic c.d.f.*** is

$$F(X_i) = \left[1 + \exp(-X_i)\right]^{-1}$$

$$= \frac{1}{\left(1 + \exp(-X_i)\right)}$$

$$= \frac{\exp(X_i)}{\left(1 + \exp(X_i)\right)}.$$

- Choice of the **standard logistic** for the distribution of the random error terms u_i leads to the ***logit* model**.

5. The Univariate Probit Model

❑ Probit Representation of the Binomial Probabilities

- In the probit model, the **binomial probabilities** $\mathbf{Pr(Y_i = 1)}$ **and** $\mathbf{Pr(Y_i = 0)}$ are represented analytically in terms of the **standard normal c.d.f.** $\Phi(Z_i)$:

$$\Phi(Z_i) = \Pr(z \le Z_i) = \int_{-\infty}^{Z_i} \phi(z)dz = \int_{-\infty}^{Z_i} (2\pi)^{-1/2} \exp\left(-\frac{z^2}{2}\right)dz$$

- The **binomial probability** $\mathbf{Pr(Y_i = 1)} = \Pr(Y_i^* > 0) = \Pr(x_i^T\beta + u_i > 0)$ is represented in the probit model as follows:

$$\begin{aligned}
\Pr(Y_i = 1) &= \Pr(Y_i^* > 0) \\
&= \Pr(x_i^T\beta + u_i > 0) \\
&= \Pr(u_i > -x_i^T\beta) \\
&= \Pr\left(\frac{u_i}{\sigma} > -\frac{x_i^T\beta}{\sigma}\right) && \text{dividing by } \sigma > 0 \\
&= 1 - \Pr\left(\frac{u_i}{\sigma} \le -\frac{x_i^T\beta}{\sigma}\right) && \text{by definition} \\
&= 1 - \Phi\left(-\frac{x_i^T\beta}{\sigma}\right) && \text{since } \frac{u_i}{\sigma} \sim N(0,1) \\
&= \Phi\left(\frac{x_i^T\beta}{\sigma}\right) && \text{by symmetry of } \phi(z) \qquad (10)
\end{aligned}$$

- The **binomial probability** $\mathbf{Pr(Y_i = 0)}$ = $\Pr(Y_i^* \le 0)$ = $\Pr(x_i^T\beta + u_i \le 0)$ is represented in the probit model as follows:

$$
\begin{aligned}
\Pr(Y_i = 0) &= \Pr(Y_i^* \le 0) \\
&= \Pr(x_i^T\beta + u_i \le 0) \\
&= \Pr(u_i \le -x_i^T\beta) \\
&= \Pr\left(\frac{u_i}{\sigma} \le -\frac{x_i^T\beta}{\sigma}\right) && \text{dividing by } \sigma > 0 \\
&= \Phi\left(-\frac{x_i^T\beta}{\sigma}\right) && \text{since } \frac{u_i}{\sigma} \sim N(0,1) \\
&= 1 - \Phi\left(\frac{x_i^T\beta}{\sigma}\right) && \text{by symmetry of } \phi(z) \qquad (11)
\end{aligned}
$$

- Note that

$$\Phi\left(\frac{x_i^T\beta}{\sigma}\right) = \Phi(Z_i) = \int_{-\infty}^{Z_i} (2\pi)^{-1/2} \exp\left(-\frac{z^2}{2}\right) dz \qquad \text{where } Z_i = \frac{x_i^T\beta}{\sigma}.$$

- The **contribution to the sample likelihood function** of the **i-th sample observation** is:

$$g(Y_i) = \left[\Phi\left(\frac{x_i^T\beta}{\sigma}\right)\right]^{Y_i}\left[1 - \Phi\left(\frac{x_i^T\beta}{\sigma}\right)\right]^{1-Y_i} \qquad Y_i = 0, 1$$

$$= \Phi\left(\frac{x_i^T\beta}{\sigma}\right) \qquad \text{for } Y_i = 1$$

$$= 1 - \Phi\left(\frac{x_i^T\beta}{\sigma}\right) \qquad \text{for } Y_i = 0$$

❑ Probit Likelihood Function

The **probit likelihood function** for a sample of N independent observations $\{Y_i : i = 1, \ldots, N\}$ is:

$$L(\beta, \sigma) = \prod_{i=1}^{N} g(Y_i)$$

$$= \prod_{i=1}^{N} \left[\Phi\left(\frac{x_i^T\beta}{\sigma}\right)\right]^{Y_i} \left[1 - \Phi\left(\frac{x_i^T\beta}{\sigma}\right)\right]^{1-Y_i} \qquad (12)$$

$$= \prod_{Y_i=1} \Phi\left(\frac{x_i^T\beta}{\sigma}\right) \prod_{Y_i=0} \left[1 - \Phi\left(\frac{x_i^T\beta}{\sigma}\right)\right]$$

❑ Probit Log-likelihood Function

- The **probit log-likelihood function** for a sample of N independent observations $\{Y_i : i = 1, \ldots, N\}$ is:

$$\begin{aligned}
\ln L(\beta, \sigma) &= \ln\left[L(\beta/\sigma)\right] \\
&= \sum_{i=1}^{N} \ln g(Y_i) \\
&= \sum_{i=1}^{N} \left\{ Y_i \ln \Phi\left(\frac{x_i^T\beta}{\sigma}\right) + (1-Y_i)\ln\left[1-\Phi\left(\frac{x_i^T\beta}{\sigma}\right)\right]\right\} \\
&= \sum_{i=1}^{N} Y_i \ln \Phi\left(\frac{x_i^T\beta}{\sigma}\right) + \sum_{i=1}^{N}(1-Y_i)\ln\left[1-\Phi\left(\frac{x_i^T\beta}{\sigma}\right)\right] \qquad (13) \\
&= \sum_{Y_i=1} \ln \Phi\left(\frac{x_i^T\beta}{\sigma}\right) + \sum_{Y_i=0}\ln\left[1-\Phi\left(\frac{x_i^T\beta}{\sigma}\right)\right]
\end{aligned}$$

- A property of the probit log-likelihood function is that the **coefficient vector β** and the **scalar parameter σ** ***are not separately identifiable***.

  Consequently, only the probit coefficient vector $\beta^* = \beta/\sigma$ can be estimated.

  However, it is conventional to impose the **normalization σ = 1**, in which case the probit coefficient vector $\beta^* = \beta$.

❑ Computing Probit Coefficient Estimates

- Maximum likelihood estimates of the probit coefficient vector β^* or β are obtained by ***maximizing*** **the probit log-likelihood function (13)** with respect to the K elements of β^* or β:

$$\text{Max}\{\beta^*\}\ \ln L(\beta^*) = \ln[L(\beta/\sigma)]$$

$$= \sum_{i=1}^{N} Y_i \ln \Phi\left(\frac{x_i^T\beta}{\sigma}\right) + \sum_{i=1}^{N}(1-Y_i)\ln\left[1-\Phi\left(\frac{x_i^T\beta}{\sigma}\right)\right]$$

$$= \sum_{i=1}^{N} Y_i \ln \Phi\left(x_i^T\beta^*\right) + \sum_{i=1}^{N}(1-Y_i)\ln\left[1-\Phi\left(x_i^T\beta^*\right)\right] \qquad (13.1)$$

where $\beta^* = \beta/\sigma$

or

$$\text{Max}\{\beta\}\ \ln L(\beta,1) = \ln[L(\beta,1)] = \sum_{i=1}^{N} Y_i \ln \Phi\left(x_i^T\beta\right) + \sum_{i=1}^{N}(1-Y_i)\ln\left[1-\Phi\left(x_i^T\beta\right)\right] \qquad (13.2)$$

- Maximization of the probit log-likelihood function (13.1)/(13.2) with respect to β^* or β requires the use of **nonlinear optimization algorithms** such as Newton's method.

- The result is an ML estimate $\hat{\beta}^* = \hat{\beta}$ of the probit coefficient vector $\beta^* = \beta$ together with an ML estimate of the covariance matrix for $\hat{\beta}^* = \hat{\beta}$, $\hat{V}(\hat{\beta}^*) = \hat{V}(\hat{\beta}) = \hat{V}_{\hat{\beta}}$.

6. The Univariate Logit Model

❑ Logit Representation of the Binomial Probabilities

- In the logit model, the **binomial probabilities** $\mathbf{Pr(Y_i = 1)}$ **and** $\mathbf{Pr(Y_i = 0)}$ are represented analytically in terms of the **standard logistic c.d.f.** $F(Z_i)$:

$$F(Z_i) = \Pr(z \le Z_i) = \frac{\exp(Z_i)}{(1 + \exp(Z_i))}.$$

- The **binomial probability** $\mathbf{Pr(Y_i = 1)} = \Pr(Y_i^* > 0) = \Pr(x_i^T\beta + u_i > 0)$ is represented in the logit model as follows:

$$\begin{aligned}
\Pr(Y_i = 1) &= \Pr(Y_i^* > 0) \\
&= \Pr(x_i^T\beta + u_i > 0) \\
&= \Pr(u_i > -x_i^T\beta) \\
&= 1 - \Pr(u_i \le -x_i^T\beta) && \text{by definition} \\
&= 1 - F(-x_i^T\beta) && \text{since } u_i \sim f(z) \\
&= F(x_i^T\beta) && \text{by symmetry of } f(z) \qquad (14)
\end{aligned}$$

- The **binomial probability** $\mathbf{Pr(Y_i = 0)} = \Pr(Y_i^* \le 0) = \Pr(x_i^T\beta + u_i \le 0)$ is represented in the logit model as follows:

$$
\begin{aligned}
\Pr(Y_i = 0) &= \Pr(Y_i^* \le 0) \\
&= \Pr(x_i^T\beta + u_i \le 0) \\
&= \Pr(u_i \le -x_i^T\beta) \\
&= F(-x_i^T\beta) && \text{by definition of } F(Z) \\
&= 1 - F(x_i^T\beta) && \text{by symmetry of } f(z) \qquad (15)
\end{aligned}
$$

- The **contribution to the sample likelihood function** of the **i-th sample observation** is:

$$g(Y_i) = \left[F(x_i^T\beta)\right]^{Y_i}\left[1 - F(x_i^T\beta)\right]^{1-Y_i} \qquad Y_i = 0, 1$$

$$= F(x_i^T\beta) \qquad \text{for } Y_i = 1$$

$$= 1 - F(x_i^T\beta) \qquad \text{for } Y_i = 0$$

❑ **Logit Likelihood Function**

The **logit likelihood function** for a sample of N independent observations $\{Y_i : i = 1, \ldots, N\}$ is:

$$L(\beta) = \prod_{i=1}^{N} g(Y_i)$$

$$= \prod_{i=1}^{N}\left[F(x_i^T\beta)\right]^{Y_i}\left[1 - F(x_i^T\beta)\right]^{1-Y_i} \qquad (16)$$

$$= \prod_{Y_i=1} F(x_i^T\beta) \prod_{Y_i=0}\left[1 - F(x_i^T\beta)\right]$$

❑ Logit Log-likelihood Function

- The **logit log-likelihood function** for a sample of N independent observations $\{Y_i : i = 1, \ldots, N\}$ is:

$$
\begin{aligned}
\ln L(\beta) &= \ln\left[L(\beta)\right] \\
&= \sum_{i=1}^{N} \ln g(Y_i) \\
&= \sum_{i=1}^{N} \ln\left\{ \left[F\left(x_i^T\beta\right)\right]^{Y_i} \left[1 - F\left(x_i^T\beta\right)\right]^{1-Y_i} \right\} \\
&= \sum_{i=1}^{N} \left\{ Y_i \ln F\left(x_i^T\beta\right) + (1 - Y_i) \ln\left[1 - F\left(x_i^T\beta\right)\right] \right\} \\
&= \sum_{i=1}^{N} Y_i \ln F\left(x_i^T\beta\right) + \sum_{i=1}^{N} (1 - Y_i) \ln\left[1 - F\left(x_i^T\beta\right)\right] \qquad (17) \\
&= \sum_{Y_i=1} \ln F\left(x_i^T\beta\right) + \sum_{Y_i=0} \ln\left[1 - F\left(x_i^T\beta\right)\right]
\end{aligned}
$$

❑ Computing Logit Coefficient Estimates by Maximum Likelihood

- Maximum likelihood estimates of the logit coefficient vector β are obtained by ***maximizing* the logit log-likelihood function (17)** with respect to the K elements of β:

$$\text{Max}\{\beta\}\ \ln L(\beta) = \ln[L(\beta)]$$

$$= \sum_{i=1}^{N} Y_i \ln F(x_i^T\beta) + \sum_{i=1}^{N}(1-Y_i)\ln[1-F(x_i^T\beta)] \qquad (17)$$

$$= \sum_{Y_i=1} \ln F(x_i^T\beta) + \sum_{Y_i=0}\ln[1-F(x_i^T\beta)]$$

- A **convenient property of the logit log-likelihood function (17)** is that it is **globally concave** with respect to the coefficient vector β.

$$\ln L(\beta) = \sum_{i=1}^{N} Y_i \ln F\left(x_i^T\beta\right) + \sum_{i=1}^{N} (1 - Y_i) \ln\left[1 - F\left(x_i^T\beta\right)\right] \quad (17)$$

$$= \sum_{Y_i=1} \ln F\left(x_i^T\beta\right) + \sum_{Y_i=0} \ln\left[1 - F\left(x_i^T\beta\right)\right]$$

This property makes nonlinear maximization of the logit log-likelihood function (17) with respect to β fairly straightforward.

The most commonly used **nonlinear optimization algorithm** for computing the ML estimates of the logit coefficients is **Newton's method**, which uses analytical first and second derivatives of $\ln L(\beta)$ with respect to β.

- The result is an ML estimate $\hat{\beta}_L$ of the logit coefficient vector β together with an ML estimate of the covariance matrix for $\hat{\beta}_L$, $\hat{V}(\hat{\beta}_L) = \hat{V}_{\hat{\beta}_L}$.