LIMITED DEPENDENT VARIABLES

Logistic Trends

One way of modelling a process of bounded growth is via a logistic function. See Figure 1. This has been used to model the growth of a population of animals in an environment with limited food resources. The simplest version of the function is

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

The second expression comes from multiplying top and bottom of the first expression by e^x.

For large negative values of x, the term $1+e^x$, in the denominator of the second expression, hardly differs from 1. Therefore, when x is negative, the logistic function resembles an exponential function.

When $x = 0$, there is is $(1 + e^x|x = 0) = 2$, and there is an inflection as the rate of increase in π begins to decline. Thereafter, the rate of increase declines rapidly toward zero, with the effect that the value of π never exceeds unity.

The inverse mapping $x = x(\pi)$ is easily derived. Consider

$$(2) \qquad \begin{aligned} 1 - \pi &= \frac{1 + e^x}{1 + e^x} - \frac{e^x}{1 + e^x} \\ &= \frac{1}{1 + e^x} = \frac{\pi}{e^x}. \end{aligned}$$

This is rearranged to give

$$(3) \qquad e^x = \frac{\pi}{1 - \pi},$$

whence the inverse function is found by taking natural logarithms:

$$(4) \qquad x(\pi) = \ln\left\{\frac{\pi}{1 - \pi}\right\}.$$

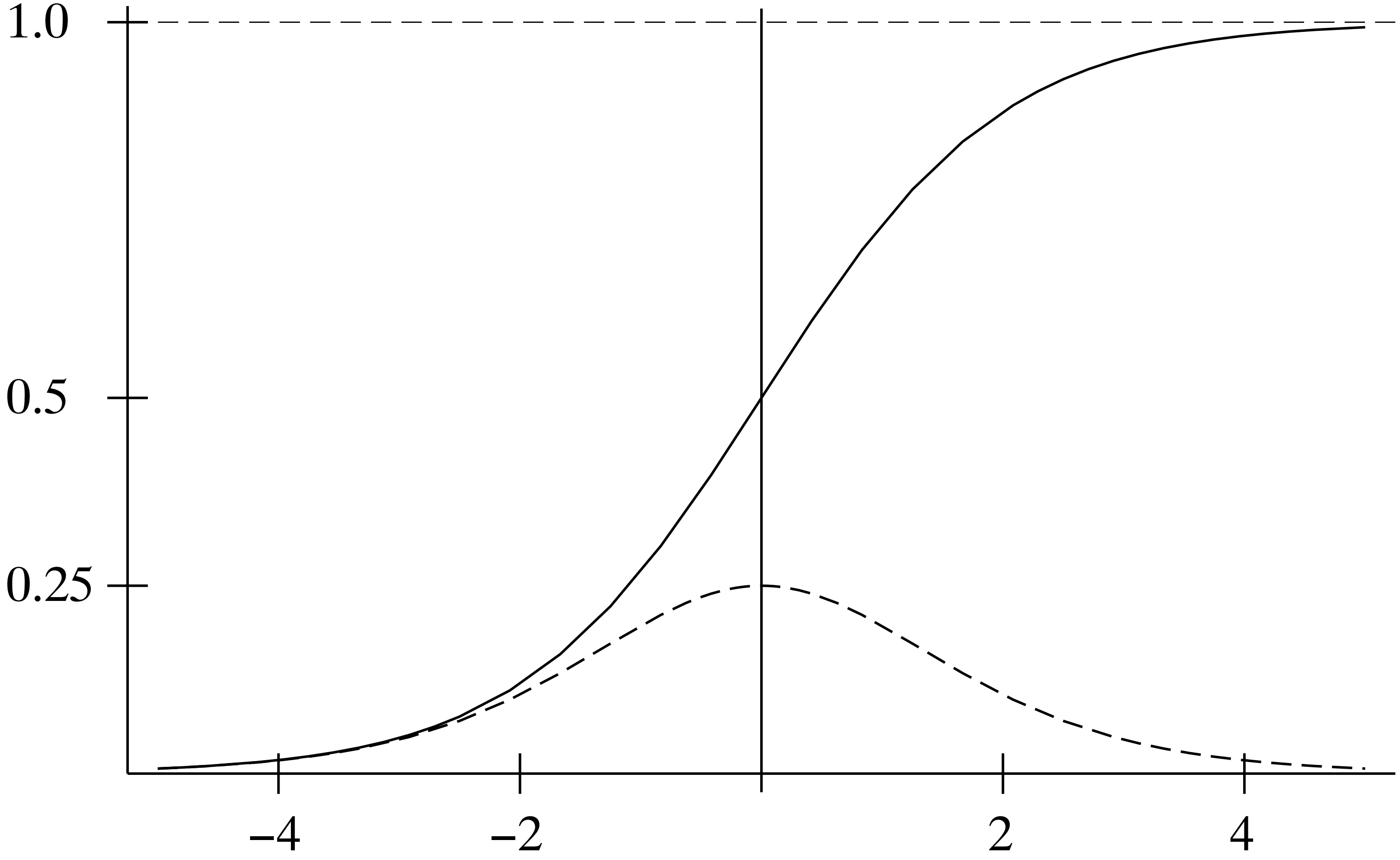

Figure 1. The logistic function $e^x/(1+e^x)$ and its derivative. For large negative values of x, the function and its derivative are close. In the case of the exponential function e^x, they coincide for all values of x.

The logistic curve needs to be elaborated before it can be fitted flexibly to a set of observations $y_1, \ldots, y_n$ tending to an upper asymptote. The general from of the function is

$$(5) \qquad y(t) = \frac{\gamma}{1 + e^{-h(t)}} = \frac{\gamma e^{h(t)}}{1 + e^{h(t)}}; \quad h(t) = \alpha + \beta t.$$

Here γ is the upper asymptote of the function, and β and α determine the rate of ascent of the function and the mid point of its ascent. It can be seen that

$$(6) \qquad \ln\left\{\frac{y(t)}{\gamma - y(t)}\right\} = h(t).$$

With the inclusion of a residual term, the equation bcomes

$$(7) \qquad \ln\left\{\frac{y_t}{\gamma - y_t}\right\} = \alpha + \beta t + e_t.$$

For a given value of γ, one may calculate the value of the dependent variable on the LHS. Then the values of α and β may be found by least-squares regression.

The value of γ may also be determined according to the criterion of minimising the sum of squares of the residuals. A crude procedure would entail running numerous regressions, each with a different value for γ. The definitive value would be the one from the regression with the least residual sum of squares.

There are other procedures for finding the minimising value of γ of a more systematic and efficient nature which might be used instead. Amongst these are the methods of Golden Section Search and Fibonnaci Search which are presented in many texts of numerical analysis.

The objection may be raised that the domain of the logistic function is the entire real line—which spans all of time from creation to eternity—whereas the sales history of a consumer durable dates only from the time when it is introduced to the market.

The problem might be overcome by replacing the time variable t in equation (15) by its logarithm and by allowing t to take only nonnegative values. See Figure 2. Then, whilst $t \in [0, \infty)$, we still have $\ln(t) \in (-\infty, \infty)$, which is the entire domain of the logistic function.

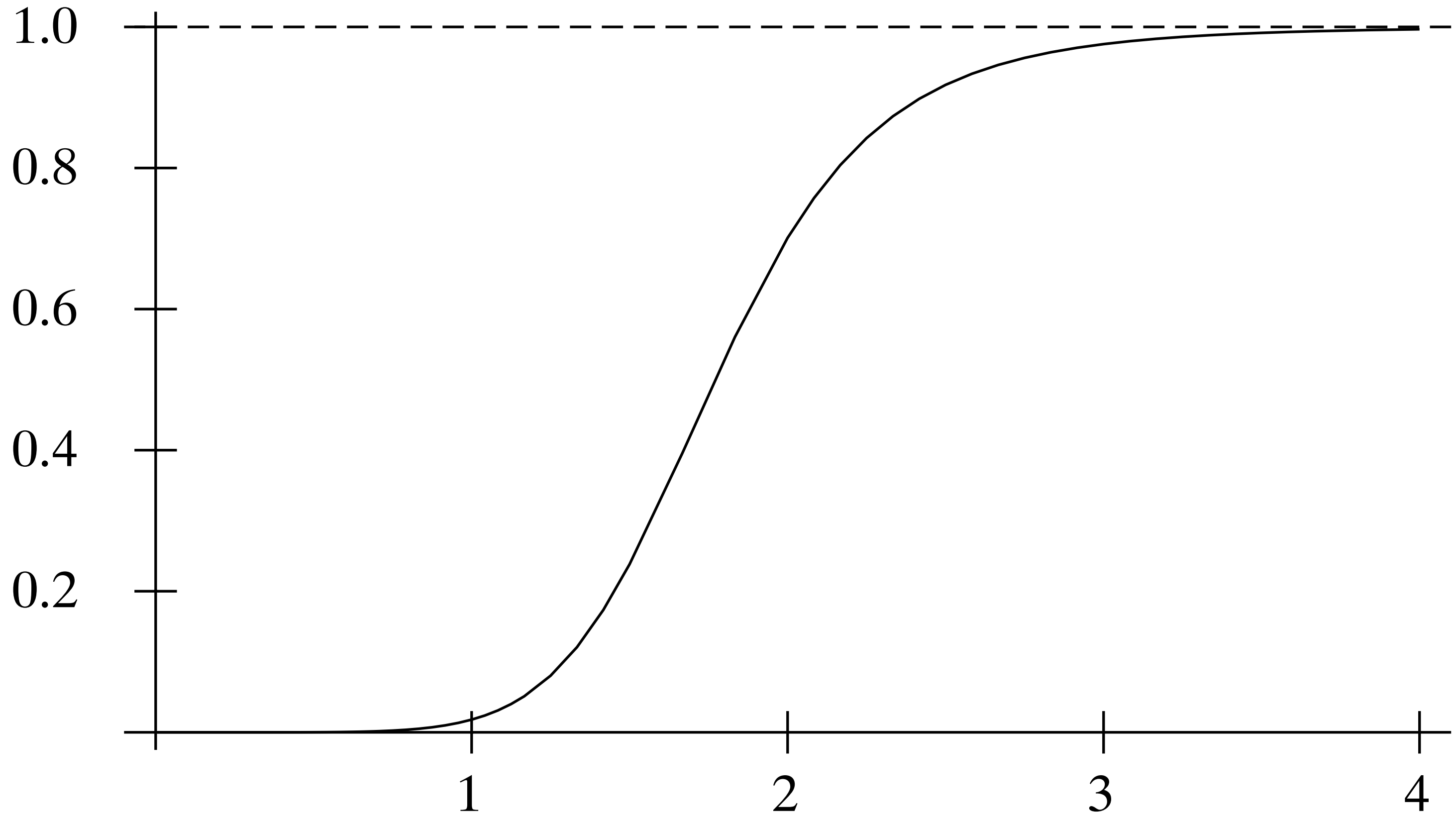

Figure 2. The function $y(t) = \gamma/(1 + \exp\{\alpha - \beta \ln(t)\})$ with $\gamma = 1$, $\alpha = 4$ and $\beta = 7$. The positive values of t are the domain of the function.

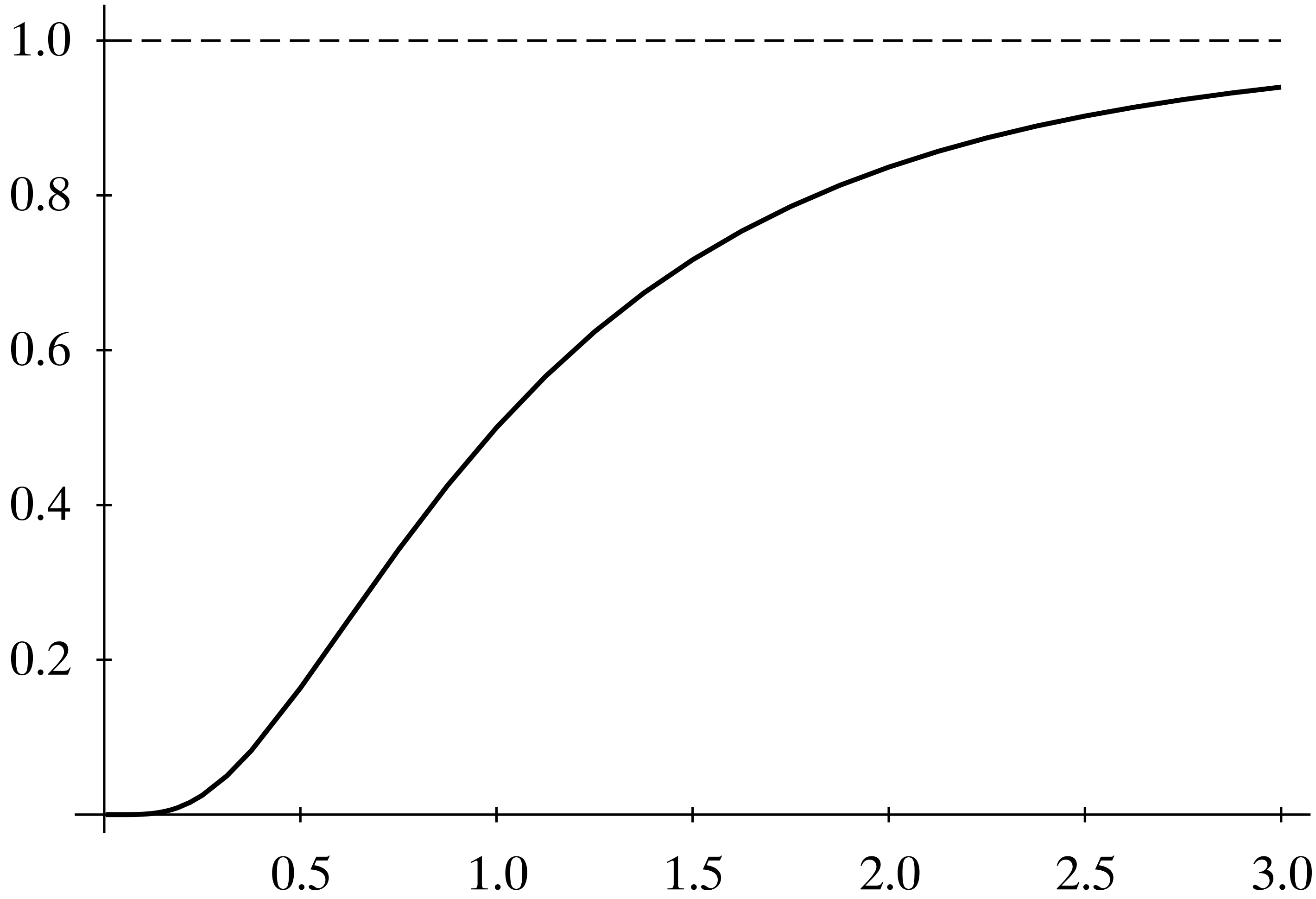

Figure 3. The cumulative log-normal distribution. The logarithm of the log-normal variate is a standard normal variate.

A Binary Dependent Variable: A Probit Model in Biology

Consider the effects of a pesticide on a sample of insects. For the ith insect, the lethal dosage is the quantity δ_i, with $\log(\delta_i) = \lambda_i \sim N(\lambda, \sigma^2)$.

If an insect is selected at random and is subjected to the dosage d_i, then the probability that it will die is $P(\lambda_i \leq x_i)$, where $x_i = \log(d_i)$. The probability is

$$(8) \qquad \pi(x_i) = \int_{-\infty}^{x_i} N(\zeta; \lambda, \sigma^2) d\zeta.$$

The function $\pi(x_i)$ with $x_i = \log(d_i)$ also indicates the fraction of the insects expected to die when all the individuals were subjected to the same global dosage $d = d_i$.

Let $y_i = 1$ if the ith insect dies and $y_i = 0$ if it survives. Then the situation of the ith insect is summarised by

$$(9) \qquad y_i = \begin{cases} 0, & \text{if } \lambda_i > x_i \quad \text{or, equivalently,} \quad \delta_i > d_i; \\ 1, & \text{if } \lambda_i \leq x_i \quad \text{or, equivalently,} \quad \delta_i \leq d_i. \end{cases}$$

The integral of (8) may be expressed in terms of a standard normal density function $N(\varepsilon; 0, 1)$. Thus

(10)
$$P(\lambda_i < x_i) \quad \text{with} \quad \lambda_i \sim N(\lambda, \sigma^2)$$

is equal to

$$P\left(\frac{\lambda_i - \lambda}{\sigma} = \varepsilon_i < h_i = \frac{x_i - \lambda}{\sigma}\right) \quad \text{with} \quad \varepsilon_i \sim N(0, 1).$$

Moreover, the standardised variable h_i, which corresponds to the dose received by the ith insect, can be written as

(11)
$$h_i = \frac{x_i - \lambda}{\sigma} = \beta_0 + \beta_1 x_i,$$

$$\text{where} \quad \beta_0 = -\frac{\lambda}{\sigma} \quad \text{and} \quad \beta_1 = \frac{1}{\sigma}.$$

To fit the model to the data, it is necessary only to estimate the parameters λ and σ^2 of the normal probability density function or, equivalently, to estimate the parameters β_0 and β_1.

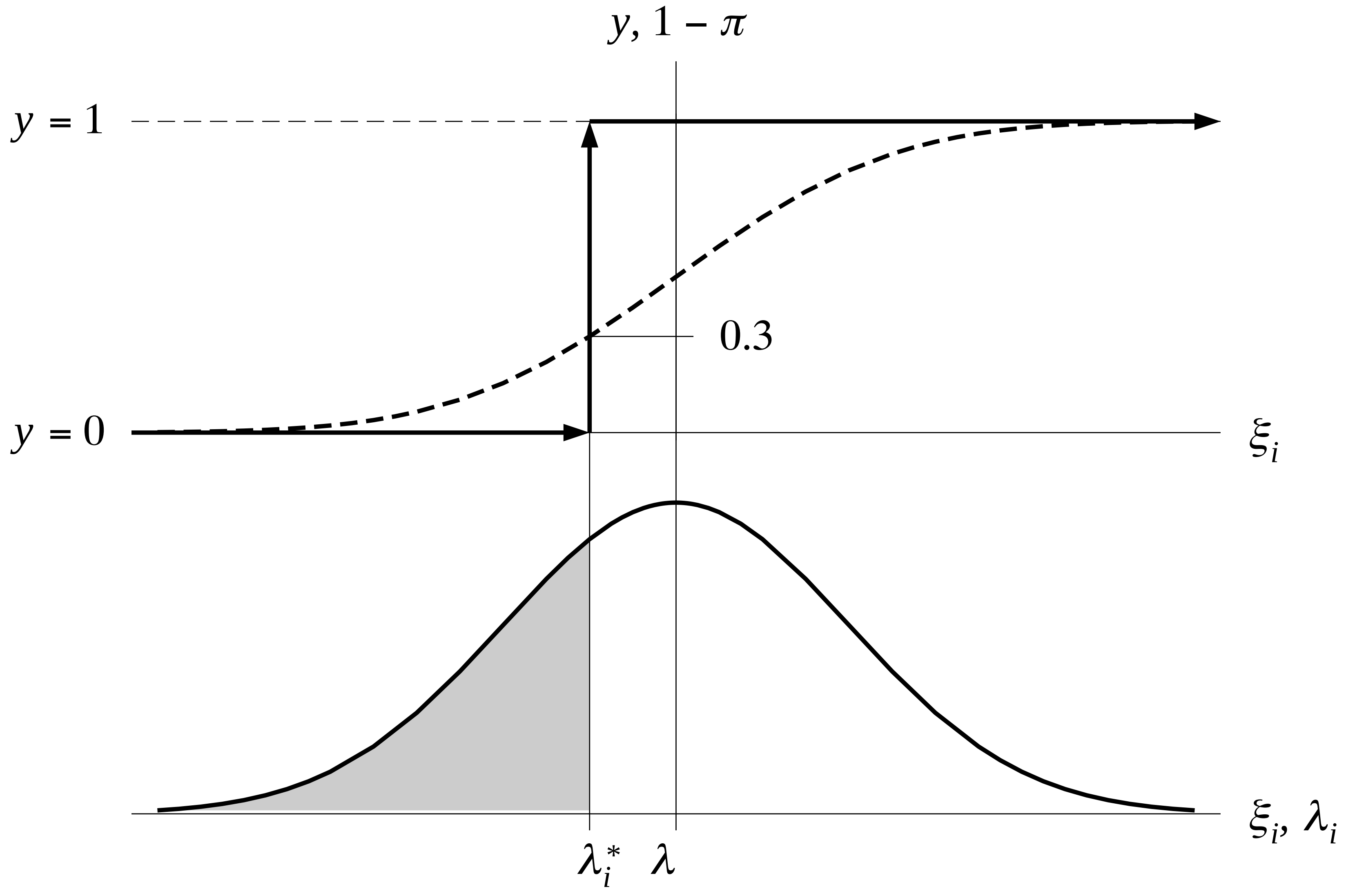

Figure 4. The probability of the threshold $\lambda_i \sim N(\lambda, \sigma^2)$ falling short of the realised value λ_i^* is the area of the shaded region in the lower diagram.

The Probit Model in Econometrics

If the stimulus ξ_i exceeds the realised threshold λ_i^*, then the step function, indicated by the arrows in the upper diagram, delivers $y = 1$. The upper diagram also shows the cumulative probability distribution function, which indicates a probability value of $P(\lambda_i < \lambda_i^*) = 1 - \pi_i = 0.3$

In econometrics, the Probit model is commonly used in describing binary choices.

The systematic influences affecting the outcome for the ith consumer may be represented by a function $\xi_i = \xi(x_{1i}, \ldots, x_{ni})$, which may be a linear combination of the variables. The idiosyncratic effects can be represented by a normal random variable of zero mean.

The ith individual will have a positive response $y_i = 1$ only if the stimulus ξ_i exceeds their own threshold value $\lambda_i \sim N(\lambda, \sigma^2)$, which is assumed to deviate at random from the level of a global threshold λ. Otherwise, there will be no response, indicated by $y_i = 0$. Thus

$$y_i = \begin{cases} 0, & \text{if } \lambda_i > \xi_i; \\ 1, & \text{if } \lambda_i \leq \xi_i. \end{cases} \tag{12}$$

These circumstances are illustrated in Figure 4.

The accompanying probability statements, expressed in term of a standard normal variate, are that

(13)

$$P(y_i = 0|\xi_i) = P\left(\frac{\lambda_i - \lambda}{\sigma} = -\varepsilon_i > \frac{\xi_i - \lambda}{\sigma}\right) \quad \text{and}$$

$$P(y_i = 1|\xi_i) = P\left(\frac{\lambda_i - \lambda}{\sigma} = -\varepsilon_i \leq \frac{\xi_i - \lambda}{\sigma}\right), \quad \text{where} \quad \varepsilon_i \sim N(0, 1).$$

On the assumption that $\xi = \xi(x_1, \ldots, x_n)$ is a linear function, these can be written as

$$\begin{aligned} &P(y_i = 0) = P\left(0 > y_i^* = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ik}\beta_k + \varepsilon_i\right) \quad \text{and} \\ &P(y_i = 1) = P\left(0 \leq y_i^* = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ik}\beta_k + \varepsilon_i\right), \end{aligned} \tag{14}$$

where

$$\beta_0 + x_{i1}\beta_1 + \cdots + x_{ik}\beta_k = \frac{\xi(x_{1i}, \ldots, x_{ki}) - \lambda}{\sigma}.$$

Thus, the original statements relating to the distribution $N(\lambda_i; \lambda, \sigma^2)$ can be converted to equivalent statements expressed in terms of the standard normal distribution $N(\varepsilon_i; 0, 1)$.

The essential quantities that require to be computed in the process of fitting the model to the data of the individual respondents, who are indexed by $i = 1, \ldots, N$, are the probability values

$$(15) \qquad P(y_i = 0) = 1 - \pi_i = \Phi(\beta_0 + x_{i1}\beta_1 + \cdots + x_{ik}\beta_k),$$

where Φ denotes the cumulative standard normal distribution function. These probability values depend on the coefficients $\beta_0, \beta_1, \ldots, \beta_k$ of the linear combination of the variables influencing the response.

Estimation with Individual Data

Imagine that we have a sample of observations $(y_i, x_{i.})$; $i = 1, \ldots, N$, where $y_i \in \{0, 1\}$ for all i. Then, assuming that the events affecting the individuals are statistically independent and taking $\pi_i = \pi(x_{i.}, \beta)$ to represent the probability that the event will affect the ith individual, we can write represent the likelihood function for the sample as

$$(16) \qquad L(\beta) = \prod_{i=1}^{N} \pi_i^{y_i}(1 - \pi_i)^{1-y_i} = \prod_{i=1}^{N} \left(\frac{\pi_i}{1 - \pi_i}\right)^{y_i} (1 - \pi_i).$$

This is the product of n point binomials. The log of the likelihood function is given by

$$\text{(17)} \qquad \log L = \sum_{i=1}^{N} y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \sum_{i=1}^{N} \log(1-\pi_i).$$

Differentiating $\log L$ with respect to β_j, which is the jth element of the parameter vector β, yields

$$\text{(18)} \qquad \begin{aligned} \frac{\partial \log L}{\partial \beta_j} &= \sum_{i=1}^{N} \frac{y_i}{\pi_i(1-\pi_i)} \frac{\partial \pi_i}{\partial \beta_j} - \sum_{i=1}^{N} \frac{1}{1-\pi_i} \frac{\partial \pi_i}{\partial \beta_j} \\ &= \sum_{i=1}^{N} \frac{y_i - \pi_i}{\pi_i(1-\pi_i)} \frac{\partial \pi_i}{\partial \beta_j}. \end{aligned}$$

To obtain the second-order derivatives which are also needed, it is helpful to write the final expression of (20) as

$$\text{(19)} \qquad \frac{\partial \log L}{\partial \beta_j} = \sum_i \left\{ \frac{y_i}{\pi_i} - \frac{1-y_i}{1-\pi_i} \right\} \frac{\partial \pi_i}{\partial \beta_j}.$$

Then it can be seen more easily that

(20)
$$\frac{\partial^2 \log L}{\partial\beta_j\beta_k} = \sum_i \left\{ \frac{y_i}{\pi_i} - \frac{1-y_i}{1-\pi_i} \right\} \frac{\partial^2 \pi_i}{\partial\beta_j\beta_k} - \sum_i \left\{ \frac{y_i}{\pi_i^2} + \frac{1-y_i}{(1-\pi_i)^2} \right\} \frac{\partial\pi_i}{\partial\beta_j}\frac{\partial\pi_i}{\partial\beta_k}.$$

The negative of the expected value of the matrix of second derivatives is the information matrix whose inverse provides the asymptotic dispersion matrix of the maximum-likelihood estimates. The expected value of the expression above is found by taking $E(y_i) = \pi_i$. On taking expectations, the first term of the RHS of (20) vanishes and the second term is simplified, with the result that

(21)
$$E\left(\frac{\partial^2 \log L}{\partial\beta_j\beta_k}\right) = \sum_i \frac{1}{\pi_i(1-\pi_i)} \frac{\partial\pi_i}{\partial\beta_j}\frac{\partial\pi_i}{\partial\beta_k}.$$

The maximum-likelihood estimates are the values which satisfy the conditions

(22)
$$\frac{\partial \log L(\beta)}{\partial\beta} = 0.$$

To solve this equation requires an iterative procedure. The Newton–Raphson procedure serves the purpose.

The Newton–Raphson Procedure

A common procedure for finding the solution or root of a nonlinear equation $\alpha(x) = 0$ is the Newton–Raphson procedure which depends upon approximating the curve $y = \alpha(x)$ by its tangent at a point near the root. Let this point be $[x_0, \alpha(x_0)]$. Then the equation of the tangent is

$$y = \alpha(x_0) + \frac{\partial \alpha(x_0)}{\partial x}(x - x_0) \tag{23}$$

and, on setting $y = 0$, we find that this line intersects the x-axis at

$$x_1 = x_0 - \left[\frac{\partial \alpha(x_0)}{\partial x}\right]^{-1} \alpha(x_0). \tag{24}$$

If x_0 is close to the root λ of the equation $\alpha(x) = 0$, then we can expect x_1 to be closer still. To find an accurate approximation to λ, we generate a sequence of approximations $\{x_0, x_1, \ldots, x_r, x_{r+1}, \ldots\}$ according to the algorithm

$$(25) \qquad x_{r+1} = x_r - \left[\frac{\partial \alpha(x_r)}{\partial x}\right]^{-1} \alpha(x_r).$$

The Newton–Raphson procedure is readily adapted to the problem of finding the value of the vector β which satisfies the equation $\partial \log L(\beta)/\partial\beta = 0$, which is the first-order condition for the maximisation of the log-likelihood function. Let β consist of two elements β_0 and β_1. Then the algorithm by which the $(r+1)$th approximation to the solution is obtained from the rth approximation is specified by

$$(26) \qquad \begin{bmatrix} \beta_0 \\ \\ \beta_1 \end{bmatrix}_{(r+1)} = \begin{bmatrix} \beta_0 \\ \\ \beta_1 \end{bmatrix}_{(r)} - \begin{bmatrix} \dfrac{\partial^2 \log L}{\partial \beta_0^2} & \dfrac{\partial^2 \log L}{\partial \beta_0 \beta_1} \\ \\ \dfrac{\partial^2 \log L}{\partial \beta_1 \beta_0} & \dfrac{\partial^2 \log L}{\partial \beta_1^2} \end{bmatrix}^{-1}_{(r)} \begin{bmatrix} \dfrac{\partial \log L}{\partial \beta_0} \\ \\ \dfrac{\partial \log L}{\partial \beta_1} \end{bmatrix}.$$

It is common to replace the matrix of second-order partial derivatives in this algorithm by its expected value which is the negative of information

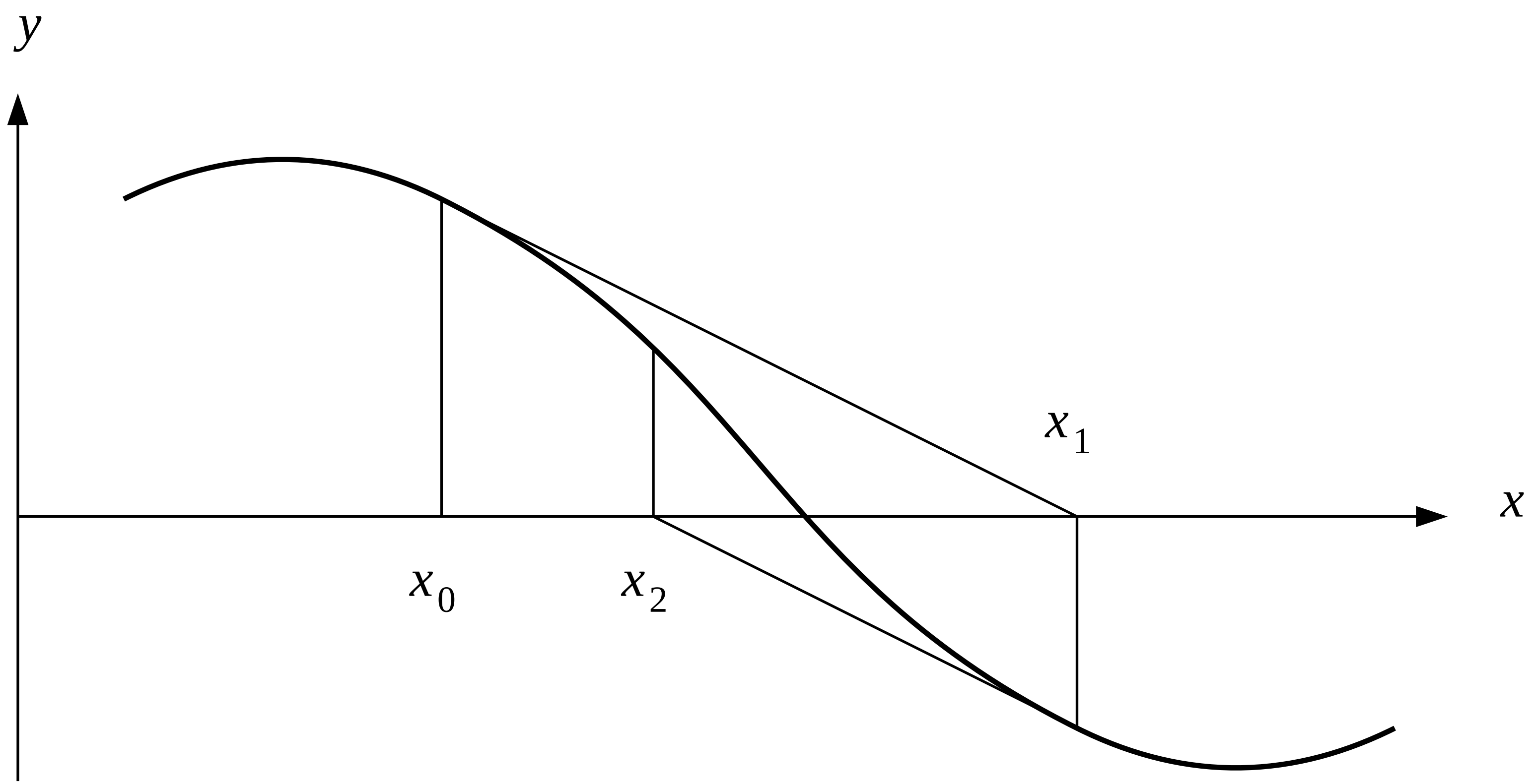

Figure 5. If x_0 is close to the root of the equation $\alpha(x) = 0$, then we can expect x_1 to be closer still.

matrix. The modified procedure is known as Fisher's method of scoring.

The algebra is often simplified by replacing the derivatives by their expectations, whereas the properties of the algorithm are hardly affected.

In the case of the simple probit model, where there is no closed-form expression for the likelihood function, the probability values, together with the various derivatives and expected derivatives to be found under (18) to (21), which are needed in order to implement one or other of these estimation procedures, may be evaluated with the help of tables which can be read into the computer.

Recall that the probability values π are specified by the cumulative normal distribution

$$\pi(h) = \int_{-\infty}^{h} \frac{1}{\sqrt{2\pi}} e^{-\zeta^2/2} d\zeta. \tag{27}$$

We may assume, for the sake of a simple illustration, that the function $h(x)$ is linear:

$$h(x) = \beta_0 + \beta_1 x. \tag{28}$$

Then the derivatives $\partial\pi_i/\partial\beta_j$ become

(29)
$$\frac{\partial\pi_i}{\partial\beta_0} = \frac{\partial\pi_i}{\partial h}.\frac{\partial h}{\partial\beta_0} = N\{h(x_i)\} \quad \text{and} \quad \frac{\partial\pi_i}{\partial\beta_1} = \frac{\partial\pi_i}{\partial h}.\frac{\partial h}{\partial\beta_1} = N\{h(x_i)\}x_i,$$

where N denotes the normal density function which is the derivative of π.

Estimation with Grouped Data

In the classical applications of probit analysis, the data was usually in the form of grouped observations. Thus, to assess the effectiveness of an insecticide, various levels of dosage $d_j; j = 1, \ldots, J$ would be administered to batches of n_j insects. The numbers $m_j = \sum_i y_{ij}$ killed in each batch would be recorded and their proportions $p_j = m_j/n_j$ would be calculated.

If a sufficiently wide range of dosages are investigated, and if the numbers n_j in the groups are large enough to allow the sample proportions p_j accurately to reflect the underlying probabilities π_j, then the plot of p_j against $x_j = \log d_j$ should give a clear impression of the underlying distribution function $\pi = \pi\{h(x)\}$.

In the case of a single experimental variable x, it would be a simple matter to infer the parameters of the function $h = \beta_0 + \beta_1 x$ from the plot.

According to the model, we have

$$\pi(h) = \pi(\beta_0 + \beta_1 x). \tag{30}$$

From the inverse $h = \pi^{-1}(\pi)$ of the function $\pi = \pi(h)$, one may obtain the values $h_j = \pi^{-1}(p_j)$. In the case of the probit model, this is a matter of referring to the table of the standard normal distribution. The values of π or p are found in the body of the table whilst the corresponding values of h are the entries in the margin. Given the points (h_j, x_j) for $j = 1, \ldots J$, it is a simple matter to fit a regression equation in the form of

$$h_j = b_0 + b_1 x_j + e_j. \tag{31}$$

In the early days of probit analysis, before the advent of the electronic computer, such fitting was often performed by eye with the help of a ruler.

To derive a more sophisticated and efficient method of estimating the parameters of the model, we may pursue a method of maximum-likelihood. This method is a straightforward generalisation of the one which we have applied to individual data.

Consider a group of n individuals which are subject to the same probability $P(y = 1) = \pi$ for the event in question. The probability that the event will occur in m out of n cases is given by the binomial formula:

$$(32) \qquad B(m, n, \pi) = \binom{n}{m} \pi^m (1 - \pi)^{n-m} = \frac{n!}{m!(n-m)!} \pi^m (1 - \pi)^{n-m}.$$

If there are J independent groups, then the joint probability of their outcomes $m_1, \ldots, m_j$ is the product

(33)

$$L = \prod_{j=1}^{J} \binom{n_j}{m_j} \pi_j^{m_j} (1 - \pi_j)^{n_j - m_j} = \prod_{j=1}^{J} \binom{n_j}{m_j} \left(\frac{\pi_j}{1 - \pi_j} \right)^{m_j} (1 - \pi_j)^{n_j}.$$

Therefore the log of the likelihood function is

$$(34) \qquad \log L = \sum_{j=1}^{J} \left\{ m_j \log \left(\frac{\pi_j}{1 - \pi_j} \right) + n_j \log(1 - \pi_j) + \log \binom{n_j}{m_j} \right\}.$$

Given that $\pi_j = \pi(x_{j.}, \beta)$, the problem is to estimate β by finding the value which satisfies the first-order condition for maximising the likelihood function which is

$$(35)\qquad \frac{\partial \log L(\beta)}{\partial \beta} = 0.$$

To provide a simple example, let us take the linear logistic model

$$(36)\qquad \pi = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}.$$

The so-called log-odds ratio is

$$(37)\qquad \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x.$$

Therefore the log-likelihood function of (34) becomes

$$(38)\quad \log L = \sum_{j=1}^{J}\left\{m_j(\beta_0+\beta_1 x_j) - n_j \log(1-e^{\beta_0+\beta_1 x_j}) + \log\binom{n_j}{m_j}\right\},$$

and its derivatives in respect of β_0 and β_1 are

(39)

$$\frac{\partial \log L}{\partial \beta_0} = \sum_j \left\{ m_j - n_j \left(\frac{e^{\beta_0+\beta_1 x_j}}{1+e^{\beta_0+\beta_1 x_j}} \right) \right\} = \sum_j (m_j - n_j\pi_j),$$

$$\frac{\partial \log L}{\partial \beta_1} = \sum_j \left\{ m_j x_j - n_j x_j \left(\frac{e^{\beta_0+\beta_1 x_j}}{1+e^{\beta_0+\beta_1 x_j}} \right) \right\} = \sum_j x_j(m_j - n_j\pi_j).$$

The information matrix, which, together with the above derivatives, is used in estimating the parameters by Fisher's method of scoring, is provided by

(40)
$$\begin{bmatrix} \sum_j m_j\pi_j(1-\pi_j) & \sum_j m_j x_j \pi_j(1-\pi_j) \\ \sum_j m_j x_j\pi_j(1-\pi_j) & \sum_j m_j x_j^2 \pi_j(1-\pi_j) \end{bmatrix}.$$