LECTURE 7

Models with Limited Dependent Variables

In this lecture, we present two models with limited dependent variables. The first is a logistic regression model in which the dependent variable has a lower bound, which is commonly taken to be zero, and an upper bound, which is approached asymptotically as the value of the independent variable, representing the systematic influences, increases. Realising such a model depends on finding a function that will map from the range of the systematic variable onto the restricted interval of the response. For this purpose, it is common to use a logistic sigmoid function.

The second model to be considered in one in which there is a binary or dichotomous response. Individuals may have to decide whether to act or not to act. Their decisions will be, to some extent, random, but they will also be affected by some measurable influences. The probability of a decision to act increases with the strength of these influences. The probability is bounded by zero and unity in a manner that is described by a cumulative probability distribution function.

Often a logistic function is used to model the probability distribution, in which case there is a so-called *logit* model. The attraction of this model is the tractability of the logistic function, which has a simple explicit form. When the distribution is normal, there is a *probit* model. The seeming disadvantage of this model is that there is no explicit form for the cumulative normal distribution function. However, the difficulty is hardly a significant one, given the availability of fast and accurate means of obtaining approximate values for the ordinates of the function. Therefore, we shall devote our attention mainly to the probit model.

Logistic Trends

Whereas unhindered exponential growth might be possible for certain monetary or financial quantities, it is implausible to suggest that such a process can be sustained for long when real resources are involved. Since real resources are finite, we expect there to be upper limits to the levels that can be attained by real economic variables, as opposed to financial variables.

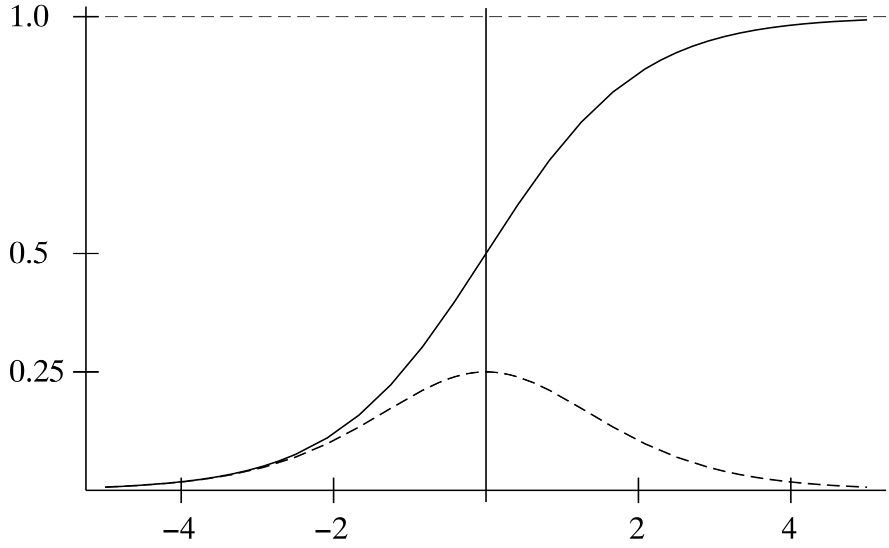

Figure 1. The logistic function $e^x/(1+e^x)$ and its derivative. For large negative values of x, the function and its derivative are close. In the case of the exponential function e^x, they coincide for all values of x.

For an example of a trend with an upper bound, we can imagine a process whereby the ownership of a consumer durable grows until the majority of households or individuals are in possession of it. Examples are provided by the sales of domestic electrical appliances such are fridges and colour television sets.

Typically, when the new durable good is introduced, the rate of sales is slow. Then, as information about the durable, or experience of it, is spread amongst consumers, the sales begin to accelerate. For a time, their cumulated total might appear to follow an exponential growth path. Then come the first signs that the market is becoming saturated; and there is a point of inflection in the cumulative curve where its second derivative—which is the rate of increase in sales per period—passes from positive to negative. Eventually, as the level of ownership approaches the saturation point, the rate of sales will decline to a constant level, which may be at zero, if the good is wholly durable, or at a small positive replacement rate if it is not.

It is very difficult to specify the dynamics of a process such as the one we have described whenever there are replacement sales to be taken into account. The reason is that the replacement sales depend not only on the size of the ownership of the durable goods but also upon the age of the stock of goods. The latter is a function, at least in an early period, of the way in which sales have grown at the outset. Often we have to be content with modelling only the growth of ownership.

One of the simplest ways of modelling the growth of ownership is to employ the so-called logistic curve. See Figure 1. This classical device has its origins in the mathematics of biology where it 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 given by

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

The second expression comes from multiplying top and bottom of the first expression by e^x. The logistic curve varies between a value of zero, which is approached as $x \to -\infty$, and a value of unity, which is approached as $x \to +\infty$. At the mid point, where $x = 0$, the value of the function is $\pi(0) = \frac{1}{2}$. These characteristics can be understood easily in reference to the first expression.

The alternative expression for the logistic curve also lends itself to an interpretation. We may begin by noting that, for large negative values of x, the term $1+e^x$, which is found in the denominator, is not significantly different from unity. Therefore, as x increases from such values towards zero, the logistic function closely resembles an exponential function. By the time x reaches zero, the denominator, with a value of 2, is already significantly affected by the term e^x. At that point, there is an inflection in the curve 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

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

This is rearranged to give

$$e^x = \frac{\pi}{1-\pi}, \tag{3}$$

whence the inverse function is found by taking natural logarithms:

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

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

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

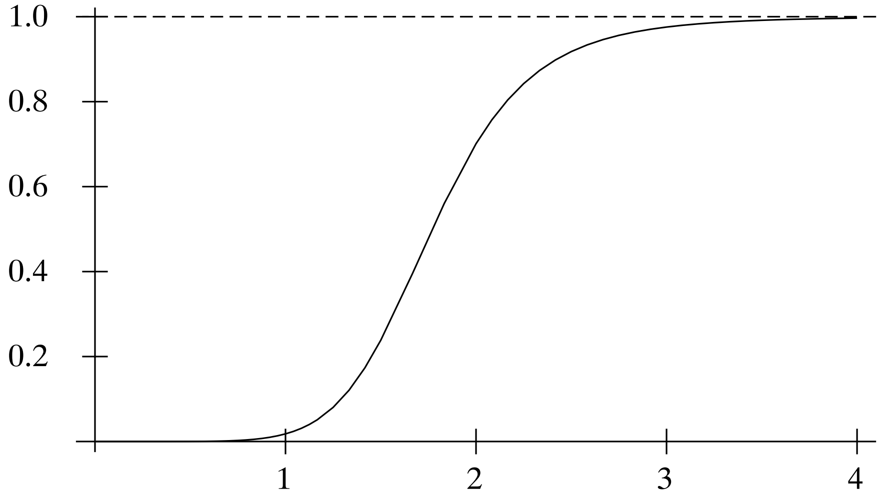

Figure 2. The function $y(t) = \gamma/(1 + \exp\{\alpha - \beta\ln(t)\})$ with $\gamma = 1$, $\alpha = 4$ and $\beta = 7$. The positive values of t are the domain of the function.

Here γ is the upper asymptote of the function, which is the saturation level of ownership in the example of the consumer durable. The parameters β and α determine respectively the rate of ascent of the function and the mid point of its ascent, measured on the time-axis.

It can be seen that

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

Therefore, with the inclusion of a residual term, the equation for the generic element of the sample is

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

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

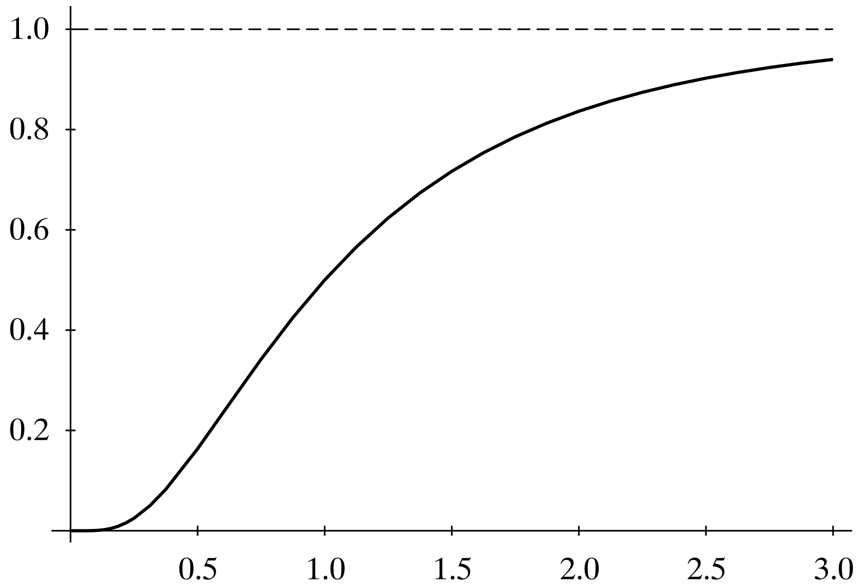

Figure 3. The cumulative log-normal distribution. The logarithm of the log-normal variate is a standard normal variate.

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.

There are many curves which will serve the purpose of modelling a sigmoidal growth process. Their number is equal, at least, to the number of theoretical probability density functions—for the corresponding (cumulative) distribution functions rise monotonically from zero to unity in ways with are suggestive of processes of bounded growth.

A Classical Probit Model in Biology

The classical example of a probit model concerns the effects of a pesticide upon a sample of insects. For the ith insect, the lethal dosage is the quantity δ_i which is the realised value of a random variable; and it is assumed that, in the population of these insects, the values $\lambda_i = \log(\delta_i)$ are distributed normally with a mean of λ and a variance of σ^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 $x_i = \log(d_i)$. This is given by

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

The function $\pi(x_i)$ with $x_i = \log(d_i)$ also indicates the fraction of a sample of insects which could be expected to die if 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 insect is summarised by writing

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

These circumstances are illustrated in Figure 4, albeit that, in the diagrams, the variable x_i has been replaced by $\xi_i = \xi(x_{1i}, \ldots, x_{ki})$, which represents the ordinate of an unspecified function of k measurable influences effecting the ith individual.

By making the assumption that it is the log of the lethal dosage which follows a normal distribution, rather than the lethal dosage itself, we avoid the unwitting implication that insects can die from negative dosages. The lethal dosages are said to have a log-normal distribution.

The log-normal distribution has an upper tail which converges rather slowly to zero, which is seen in Figure 3. Therefore, the corresponding tail of the cumulative distribution converges slowly to the upper asymptote of unity, which implies that some individuals are virtually immune to the effects of the pesticide. In a laboratory experiment, one would expect to find, to the contrary, that there is a moderate dosage that is certain to kill all the insects. In the field, however, there is always the chance that some insects will be sheltered from the pesticide.

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

$$(10) \qquad \begin{aligned} & P(\lambda_i < x_i) \quad \text{with} \quad \lambda_i \sim N(\lambda, \sigma^2) \\ & \text{is equal to} \\ & P\left(\frac{\lambda_i - \lambda}{\sigma} = \varepsilon_i < h_i = \frac{x_i - \lambda}{\sigma}\right) \quad \text{with} \quad \varepsilon_i \sim N(0, 1). \end{aligned}$$

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

$$(11) \qquad \begin{aligned} & h_i = \frac{x_i - \lambda}{\sigma} = \beta_0 + \beta_1 x_i, \\ & \text{where} \quad \beta_0 = -\frac{\lambda}{\sigma} \quad \text{and} \quad \beta_1 = \frac{1}{\sigma}. \end{aligned}$$

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.

The Probit Model in Econometrics

In econometrics, the Probit model is commonly used in describing binary choices. The circumstances of these choices are not the life-threatening ones

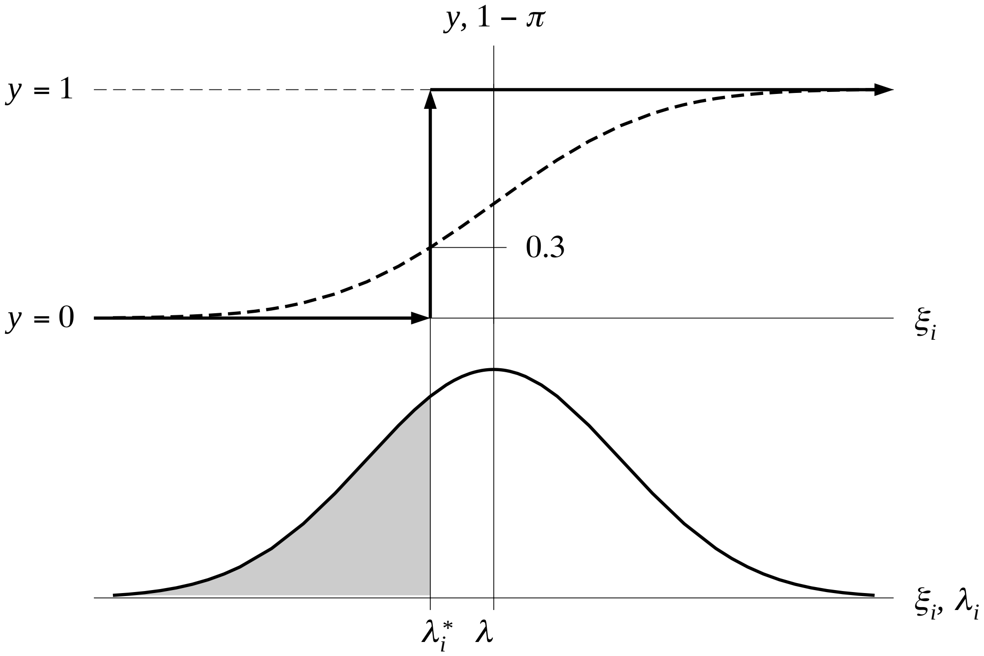

Figure 4. The probability of the threshold $\lambda_i \sim N(\lambda, \sigma^2)$ falling short of the realised value λ_i^* is the area of the shaded region in the lower diagram. If the stimulus ξ_i exceeds the realised threshold λ_i^*, then the step function, indicated by the arrows in the upper diagram, delivers $y = 1$. The upper diagram also shows the cumulative probability distribution function, which indicates a probability value of $P(\lambda_i < \lambda_i^*) = 1 - \pi_i = 0.3$.

that affect the insects; and the issue is typically a matter of whether or not a consumer will purchase an item of a particular good or whether or not they will decide to pursue a particular activity. There may be numerous influences affecting the outcome; and these will include the consumer's own idiosyncratic departure from the mean level of susceptibility, denoted by λ.

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

$$(13)\qquad \begin{aligned} P(y_i = 0|\xi_i) &= P\left(\frac{\lambda_i - \lambda}{\sigma} = -\varepsilon_i > \frac{\xi_i - \lambda}{\sigma}\right) \quad \text{and} \\ P(y_i = 1|\xi_i) &= P\left(\frac{\lambda_i - \lambda}{\sigma} = -\varepsilon_i \leq \frac{\xi_i - \lambda}{\sigma}\right), \quad \text{where} \quad \varepsilon_i \sim N(0,1). \end{aligned}$$

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

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

where

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

This is a common formulation. Thus, by employing the appropriate normalising transformations, it is possible to convert the original statements relating to the normal distribution $N(\lambda_i; \lambda, \sigma^2)$ 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

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

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

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

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

$$\frac{\partial \log L}{\partial \beta_j} = \sum_i \left\{ \frac{y_i}{\pi_i} - \frac{1-y_i}{1-\pi_i} \right\} \frac{\partial \pi_i}{\partial \beta_j}. \tag{19}$$

Then it can be seen more easily that

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

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

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

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

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

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

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

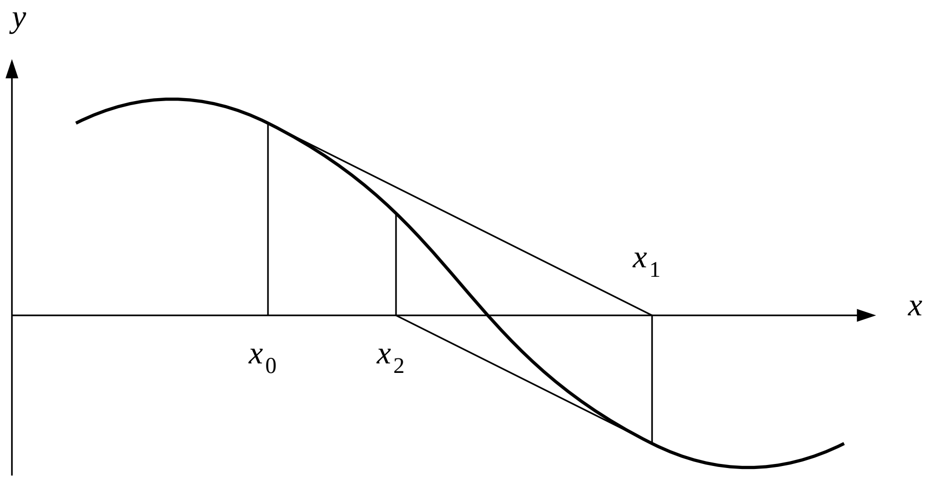

Figure 5. If x_0 is close to the root of the equation $\alpha(x) = 0$, then we can expect x_1 to be closer still.

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

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

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

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

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 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

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

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:

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

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

$$L = \prod_{j=1}^{J} \binom{n_j}{m_j} \pi_j^{m_j} (1-\pi_j)^{n_j - m_j} = \prod_{j=1}^{J} \binom{n_j}{m_j} \left(\frac{\pi_j}{1-\pi_j} \right)^{m_j} (1-\pi_j)^{n_j}. \tag{33}$$

Therefore the log of the likelihood function is

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

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

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

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

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

The so-called log-odds ratio is

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

Therefore the log-likelihood function of (34) becomes

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

and its derivatives in respect of β_0 and β_1 are

$$\begin{aligned} \frac{\partial \log L}{\partial \beta_0} &= \sum_j \left\{ m_j - n_j \left(\frac{e^{\beta_0 + \beta_1 x_j}}{1 + e^{\beta_0 + \beta_1 x_j}} \right) \right\} = \sum_j (m_j - n_j \pi_j), \\ \frac{\partial \log L}{\partial \beta_1} &= \sum_j \left\{ m_j x_j - n_j x_j \left(\frac{e^{\beta_0 + \beta_1 x_j}}{1 + e^{\beta_0 + \beta_1 x_j}} \right) \right\} = \sum_j x_j (m_j - n_j \pi_j). \end{aligned} \tag{39}$$

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

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