第六章　発展した分析法[1]

劉慶豊[2]

小樽商科大学

July 11, 2011

[1]第二章からの資料は森棟公夫先生著「基礎コース　計量経済学」をもとに作成したものである。

[2]E-mail:qliu@res.otaru-uc.ac.jp, URL:http://www.otaru-uc.ac.jp/~qliu/

離散選択モデル

- 被説明変数 y_i が $0, 1, 2$ などの離散的な値を取るようなモデルを考える
- 例：女性が就職するなら 1、専業主婦になるなら 0。通勤手段として、車なら 0、電車なら 1、バスなら 2。
- 離散的な変数で表す選択肢を選択する原因を分析する。
- 二項と多項選択モデルがある。

二項選択モデル

- y_i が 0 と 1 しか取らない。例：持ち家を持つなら 1、持たないなら 0。
- 0 と 1 をどっちを選ぶかの確率を

$$P(y_i = 1) = p_i$$
$$P(y_i = 0) = 1 - p_i = q_i$$

  と表す。
- 確率 p_i、q_i のモデル化の違いによって線形モデル、プロビットモデルとロジットモデルがある。

分析例

- 給与所得者が家を所有するか否かの確率を分析する
- 被説明変数：家を持つか持たないか。持ち家を持つなら 1、持たないなら 0。
- 説明変数：所得、勤続年数、年齢、家族構成など

一般表現

- **モデル式の一般表現**

$$P(y_i = 1) = p_i$$
$$P(y_i = 0) = 1 - p_i$$

$$y_i = p_i + \varepsilon_i \tag{1}$$

$$\begin{aligned} p_i &= f\left(x_{1i}, x_{2i}, \cdots, x_{Ki}\right) \\ &= f\,\textbf{(所得}_i\textbf{、勤続年数}_i\textbf{、年齢}_i\textbf{、}\cdots\textbf{)} \end{aligned}$$

線形確率モデル

- **線形確率モデルのモデル式**

$$
\begin{aligned}
p_i &= f\left(x_{1i}, x_{2i}, \cdots, x_{Ki}\right) \\
&= \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki}
\end{aligned}
$$

線形確率モデルの問題点 p_i **が確率であるため、0と1の間の値を取るべきだが、**$\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki}$ **が必ずしも0と1の間の値にならないため、線形確率モデルはこのことを保障できない。**

プロビットモデル

- **プロビットモデルのモデル式**

$$p_i = \int_{-\infty}^{f_i} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)dt$$

ただし、

$$f_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki}$$

ロジットモデル

- **ロジットモデルのモデル式**

$$p_i = \frac{\exp(f_i)}{1 + \exp(f_i)}$$

ただし、

$$f_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki}$$

最尤推定法

- 最尤推定法でプロビットモデルとロジットモデルを推定する

尤度関数 尤度関数は観測された被説明変数のデータの同時確率である。

- 尤度関数を求めてみよう。
  1. $(p_i)^{y_i}(1-p_i)^{1-y_i}$ は i 番目の観測値が y_i になる確率である。(確認してみる：$y_i = 1$ の確率を $(p_i)^{y_i}(1-p_i)^{1-y_i}$ で計算すると $(p_i)^1(1-p_i)^{1-1} = p_i$。これはモデル一般表現のページにある定義による $P(y_i = 1) = p_i$ と一致する。$y_i = 0$ の確率に関しても同様に確認できる。)
  2. $y_1, y_2, \cdots, y_n$ が互いに独立であるとする。$y_1, y_2, \cdots, y_n$ 同時確率はそれぞれの確率（周辺確率）の掛け算となる。
  3. 従って、尤度関数は観測された被説明変数のデータの同時確率であるため、尤度関数

$$L = \prod_{i=1}^{n}(p_i)^{y_i}(1-p_i)^{1-y_i}$$

となる。$\prod_{i=1}^{n}$ は積和を表す。

最尤推定法（続き）

対数尤度関数 尤度関数の対数である

$$\ln(L) = \sum_{i=1}^{n}\{y_i \ln(p_i) + (1 - y_i)\ln(1 - p_i)\}$$

最尤推定法 最尤推定法はプロビットモデルやロジットモデルの p_i の定義を対数尤度関数に代入して、対数尤度関数を最大にできるような $\beta_1, \beta_2, \cdots, \beta_K$ の値を係数の推定値とする方法である。（対数尤度関数を最大にすることは尤度関数を最大にすることと同等である。）

直感的な理解 最尤推定法は $y_1, y_2, \cdots, y_n$ の同時確率分布を推定しているような方法である。その確率分布の元では「今のデータが一番現れやすい」となるような確率分布を探していることとなる。

プロビットモデルとロジットモデルのファイナンスや保険への応用

倒産リスクへの応用 i番目の企業は倒産したら$y_i = 1$倒産していないなら$y_i = 0$とする。説明変数としてその企業の売り上げや資本金や収益率などの財務指標が考えられる。

保険への応用 i番の種類の幾つかの特性を持つ人が一年以内に入院する確率を分析する。入院するなら$y_i = 1$、していないなら$y_i = 0$とする。説明変数としてi番の種類の特性が考えられる。