質的従属変数のロジット、プロビット分析1†

担当：　長倉　大輔
（ながくらだいすけ）

† この資料は私のゼミにおいて使用するために作成した資料です。ゼミのWEBページ上で公開しており、自由に参照して頂いて構いません。ただし、内容について、一応検証してありますが、もし間違いがあった場合でもそれによって生じるいかなる損害、不利益について責任を負いかねますのでご了承ください。間違いは発見次第、継続的に直していますが、まだ存在する可能性があります。

ロジット、プロビット分析

- ロジット、プロビット分析

ロジット、プロビット分析とは分析対象が、数値データではない時に用いられる。

例えば

「車を購入するかどうかの決定要因の分析」
「主婦がパートで働くかどうかの決定要因の分析」
「結婚するかどうかの決定要因の分析」

などである。

ロジット、プロビット分析

- 質的従属変数

例えば車の購入の決定要因の分析であれば、観測されるのは n 人の個人に対して「各個人が車を購入したか否か?」である。これは数値データではない。

ただし、この場合も i 番目の個人(これを個人 i と呼ぼう)が車を購入したら $Y_i = 1$、購入しなかったら $Y_i = 0$ となるような変数 $Y_i, i = 1, \ldots, n$ を定義し、人工的に数値データを作る事ができる。

このような変数を従属変数として使った場合、**質的従属変数**と呼ばれる。

ロジット、プロビット分析

- 確率の推定

このような質的従属変数に対してロジット、プロビットモデルによって、「$Y_i = 1$となる**確率**にはどのような要因が影響を与えるか？」を分析することができる。

ではそのような確率と、決定要因となりうる変数との関係をどのように定式化したらよいのだろうか？

よく使われるモデルは**潜在変数モデル**と呼ばれるものである。

ロジット、プロビット分析

- 潜在変数モデル

潜在変数モデルとは選択において、個人 i はそれぞれの**選択肢から得られる効用を比較する**事によって、選択を行うというモデルである。以下では「車の購入」を例にとって説明する。

個人 i が車を購入することによって得られる効用を U_{i1} 、車を購入しないことによって得られる効用を U_{i0} としよう。

個人 i は $U_{i1} > U_{i0}$ の時、車を購入するとする
(この時 $Y_i = 1$ である)。

ロジット、プロビット分析

- 潜在変数モデル

効用 U_{i0}, U_{i1} の定式化として、効用に影響を与えそうな要因の線形の関数がよく用いられる。例えば、車の購入の例では

$$U_{i1} = \alpha_1 + \beta_1 X_{i1} + \kappa_1 X_{i2} + \varphi_1 X_{i3} + \varepsilon_{i1}$$
$$U_{i0} = \alpha_0 + \beta_0 X_{i1} + \kappa_0 X_{i2} + \varphi_0 X_{i3} + \varepsilon_{i0}$$

X_{i1} : 個人 i の年収, X_{i2} : 車の平均価格,
X_{i3} : 地域のガソリン価格,

などのように定式化される。ここで $\alpha_k, \beta_k, \kappa_k, \varphi_k$, は定数、$\varepsilon_{i0}, \varepsilon_{i1}$ はこれらの要因だけでは説明できない部分をまとめたものである。

ロジット、プロビット分析

- 潜在変数モデル

(以後、説明の簡単化のため説明変数は X_{i1} だけとする）
この時、$\Pr(Y_i = 1)$ は

$$
\begin{aligned}
&\Pr(U_{i0} < U_{i1}) \\
&= \Pr(\alpha_0 + \beta_0 X_{i1} + \varepsilon_{i0} < \alpha_1 + \beta_1 X_{i1} + \varepsilon_{i1}) \\
&= \Pr(\varepsilon_{i0} - \varepsilon_{i1} < \alpha_1 - \alpha_0 + (\beta_1 - \beta_0) X_{i1}) \\
&= \Pr(\eta_i < \gamma + \delta X_{i1})
\end{aligned}
$$

ここで $\eta_i = \varepsilon_{i0} - \varepsilon_{i1}$, $\gamma = \alpha_1 - \alpha_0$, $\delta = \beta_1 - \beta_0$ である。

ロジット、プロビット分析

- 潜在変数モデル

η_i は確率的に変動し、それぞれの i について共通の(連続型の)確率分布に従うとみなせば

$$\Pr(\eta_i < \gamma + \delta X_{i1}) = F_\eta(\gamma + \delta X_{i1})$$

と表す事ができる。ここで $F_\eta(x)$ はこの確率変数 η_i の分布関数である。よって

$$\Pr(Y_i = 1) = F_\eta(\gamma + \delta X_{i1})$$

となる。ここで γ, δ が推定するパラメータである。

η_i に **ロジスティック分布**を仮定すれば、**ロジットモデル**、**標準正規分布**を仮定すれば、**プロビットモデル**と呼ばれる

ロジット、プロビット分析

- ロジスティック分布の分布関数

ロジスティック分布の分布関数は

$$F_\eta(x) = \Lambda(x) = \frac{\exp(x)}{1+\exp(x)}$$

で与えられる。

- 標準正規分布の分布関数

正規分布の分布関数は明示的に求められない。しかしさまざまなよい近似法が存在し、それによって計算する。
標準正規分布の分布関数は通常 $\Phi(x)$ で表される。

ロジット、プロビット分析

- ロジット、プロビットモデルの最尤推定

Y_i の従う確率分布が分かったので、同時確率関数より尤度関数を構成することができ、それを最大化することによってパラメーター γ, δ を推定することができる。

$F_i = F_\eta(\gamma + \delta X_{i1})$ とすると、Y_i に関して $\Pr(Y_i = 1) = F_i$, $\Pr(Y_i = 0) = 1 - F_i$ である。これをまとめると

$$\Pr(Y_i = y) = F_i^y (1 - F_i)^{1-y}, \quad y = 0, 1$$

と書ける。

ロジット、プロビット分析

- ロジット、プロビットモデルの最尤推定

よって η_i が独立だとすると(これは仮定する)、
$Y_i = y_i$, ($y_i = 0$ か 1) が観察された時の尤度は

$$
\begin{aligned}
&L(\boldsymbol{\theta}) \\
&= \Pr(Y_1 = y_1) \times \Pr(Y_2 = y_2) \times \cdots \times \Pr(Y_n = y_n) \\
&= F_1^{y_1}(1 - F_1)^{1-y_1} \times F_2^{y_2}(1 - F_2)^{1-y_2} \times \cdots \times F_n^{y_n}(1 - F_n)^{1-y_n} \\
&= \prod_{i=1}^{n} F_i^{y_i}(1 - F_i)^{1-y_i}
\end{aligned}
$$

となる。ここで$\boldsymbol{\theta} = (\gamma, \delta)$ である。

ロジット、プロビット分析

- ロジット、プロビットモデルの最尤推定

対数尤度関数は

$$\log L(\boldsymbol{\theta}) = \log(\prod_{i=1}^{n} F_i^{y_i}(1-F_i)^{1-y_i})$$

$$= \sum_{i=1}^{n} y_i \log F_i + \sum_{i=1}^{n} (1-y_i)\log(1-F_i)$$

となる。これを最大化するような $\boldsymbol{\theta}$ の値が最尤推定値である。

ロジット、プロビット分析

- 最尤推定量の分布

この最尤推定量は一致性をもち、**漸近的に正規分布に従う**。よってその標準誤差(つまり標準偏差)が推定できれば、パラメーターについての仮説検定を行う事ができる。

ロジット、プロビット分析

- パラメーターについての仮説検定

任意の i 番目のパラメーター θ_i について、帰無仮説

$$H_0: \theta_i = \theta_0$$

を検定するためには

$$Z_T = \frac{\hat{\theta}_{i,ML} - \theta_0}{\sigma(\hat{\theta}_{i,ML})}$$

というような統計量を用いればよい。ここで $\hat{\theta}_{i,ML}$ は θ_i の最尤推定量、$\sigma(\hat{\theta}_{i,ML})$ は $\hat{\theta}_{i,ML}$ の標準誤差である。

Z_T は帰無仮説の下で漸近的に**標準正規分布**に従う。

ロジット、プロビット分析

- 限界確率効果(Marginal Probability Effect)

個人 i の j 番目変数 X_{ij} の**限界確率効果** MPE_{ij} は

$$MPE_{ij} = \frac{\partial \Pr(Y_i = 1)}{\partial X_{ij}}$$

と定義される。これは説明変数 X_{ij} が(わずかに) 変化した時に $Y_i = 1$ となる確率 $\Pr(Y_i = 1)$ がどの程度変化するかを表している。

MPE_{ij} の大きい変数ほど、個人 i の行動により大きな影響を与える変数であると解釈できる。

ロジット、プロビット分析

- 限界確率効果(Marginal Probability Effect)

$\Pr(Y_i = 1) = F_\eta(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})$, $\mathbf{X}_i = [X_{i1},\ldots, X_{iK}]^{\mathrm{T}}$, $\boldsymbol{\beta} = [\beta_1,\ldots, \beta_K]^{\mathrm{T}}$ とすると、限界確率効果は $\Pr(Y_{\mathrm{i}} =1)$ を X_{ij} で微分したものであるので、合成関数の微分の法則より

$$MPE_{ij} = \frac{\partial F_\eta(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})}{\partial X_{ij}} = f_\eta(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})\beta_j$$

で与えられる。ここで $f_\eta(x) = \partial F_\eta(x)/\, \partial x$ である。

ロジット、プロビット分析

- ロジット、プロビットモデルの限界確率効果

プロビットモデルにおいて $F_\eta(x) = \Phi(x)$ であり、$f_\eta(x) = \partial F_\eta(x)/\partial x = \phi(x)$ である。ここで $\phi(x)$ は**標準正規分布の密度関数を表す**。よって

$$MPE_{ij} = \phi(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})\beta_j$$

である。またロジットモデルにおいては $F_\eta(x) = \Lambda(x)$ であるので

$$MPE_{ij} = \Lambda(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})[1-\Lambda(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})]\beta_j$$

となる(演習問題)。ここで $\Lambda(x) = \dfrac{\exp(x)}{1+\exp(x)}$ である。

ロジット、プロビット分析

- 平均限界効果 (Average Marginal Probability Effect)

限界確率効果は個人 i について、変数 X_{ij} の(実現値からの微小な変化の)選択行動への影響を見るものであった。これに対して、全個人の限界確率効果の平均をとったものを**平均限界効果**と呼び

$$AMPE_{ij} = \frac{1}{n}\sum_{i=1}^{n} f_{\eta}(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})\beta_j$$

と定義される。これは全体的な傾向を見るのに適している。

ロジット、プロビット分析

- 期待限界効果 (Expected Marginal Probability Effect)

また、限界確率効果を $X_{ij},\ i=1,\ldots,n,\ j=1,\ldots,K$ の i についての平均値 $\overline{\mathbf{X}}=[\overline{X}_1,\overline{X}_2,\ldots,\overline{X}_K]^{\mathrm{T}},\ \overline{X}_j=\sum_{i=1}^{n}X_{ij}$, $j=1,\ldots,K$ で評価したものを**期待限界効果**と呼び

$$EMPE_{ij}=f_\eta(\overline{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{\beta})\beta_j$$

と定義される。これも全体の傾向を総合的に見るのに適している。

演習問題

問題1 ロジスティック分布の密度関数を求めなさい。これより、ロジットモデルにおいて MPE_{ij} が前述のように与えられることを示しなさい。

問題2 今、個人i $(i = 1,\ldots,3)$ はエアコンを買うかどうか迷っている。個人i がエアコンを買うことから得る効用を U_{i1}, 買わないことによる効用を U_{i0} としよう。U_{ij} $(j = 0, 1)$ はある説明変数 X_i (例えば住んでいる地域の夏の平均気温) との間に $U_{ij} = \alpha_j + \beta_j X_i + \varepsilon_{ij}$ という関係があるとする。ここで ε_{ij} は X_i だけでは説明しきれない個人i の効用であり、確率的に変動するとする。個人i は$U_{i1} > U_{i0}$ であればエアコンを購入し、そうでなければ購入しないとする。$\eta_i = \varepsilon_{i0} - \varepsilon_{i1}$, $\gamma = \alpha_1 - \alpha_0$, $\delta = \beta_1 - \beta_0$、として$\eta_i$ はロジスティック分布に従うとする。

演習問題

(1) 個人iがエアコンを購入する確率を明示的に表しなさい。

(2) 個人iがエアコンを購入しない確率を明示的に表しなさい。

(3) Y_iを個人iがエアコンを購入したらY_i=1、購入しなかったら$Y_i = 0$を取る確率変数だとする。今$Y_1 = 0$, $Y_2 = 1$, $Y_3 = 1$が観測されたとする。この時γとδの最尤推定値を求める対数尤度関数を求めなさい。