質的従属変数の
ロジット、プロビット分析3†

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

- 順序ロジット(プロビット)モデル

複数の選択肢からの個人の選択を扱うモデルとして、多項ロジットモデルの他に**順序ロジット(プロビット)モデル**と呼ばれるモデルもよく使われる。これは選択肢に順序があるような場合のモデルである。

- (例)　(アンケート)

例えば、何らかの製品やサービスに対してのアンケートの回答として「大変良い」「良い」「普通」「悪い」「大変悪い」の一つを選ぶようなものがあるが、このような場合この5つの選択肢には明らかな順序関係が存在する。

ロジット、プロビット分析

- 順序ロジット(プロビット)モデルの定式化

(説明の簡単化のためにまず選択肢が3つの場合を考える)　今、j =1, 2, 3 の3個の選択肢があるとし、これらの選択肢には 1, 2, 3 と並べられるような何らかの自然な順序関係があるとする(例えば 選択肢1は「悪い」、選択肢2は「普通」、選択肢3は「良い」を表すなど) 。

個人 i の特性を表す説明変数を X_i としよう(簡単化のために1つだけとする)。この時、個人 i は**以下の基準に沿って**この3つの選択肢から1つを選ぶとする。

ロジット、プロビット分析

- 順序ロジット(プロビット)モデルの定式化

個人i は

$$
\begin{aligned}
& \beta X_i + \varepsilon_i < c_1 \quad \text{であれば　選択肢 1 を} \\
c_1 \le\ & \beta X_i + \varepsilon_i < c_2 \quad \text{であれば　選択肢2を} \\
c_2 \le\ & \beta X_i + \varepsilon_i \quad \text{であれば　選択肢3を}
\end{aligned}
$$

選ぶとする。ここで c_1, c_2 はある定数、β はすべての個人に共通の係数、ε_i は個人 i の選好について βX_i だけでは説明しきれない部分をまとめたものである。

ε_i の分布としてロジスティック分布を仮定したものを**順序ロジットモデル**、標準正規分布を仮定したものを**順序プロビットモデル**という。

ロジット、プロビット分析

- 順序ロジット(プロビット)モデルの定式化

例えば先ほどの商品のアンケートの例では

$$U_i = \beta X_i + \varepsilon_i$$

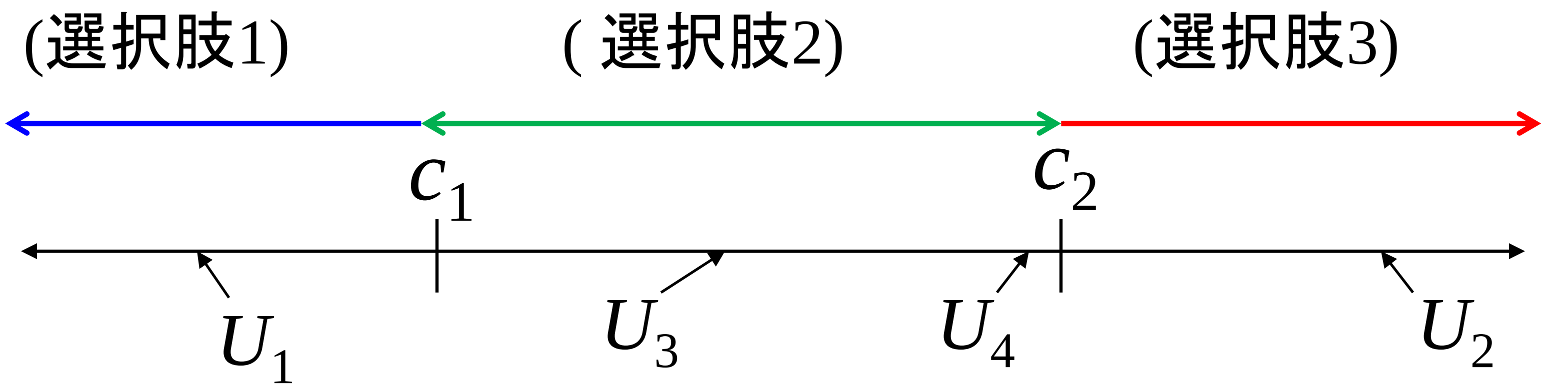
を個人 i の商品に対する満足度を示すものと考えることができる。その商品から一定以上の満足度を得れば「良い」を、ある範囲の満足度であれば「普通」を、一定以下の満足度であれば「悪い」を選択すると解釈できる。

例えばこの「商品」を「風邪薬」とすると、X_i としては個人 i が「薬を服用後、風邪が治るまでの時間」などとできるであろう。

ロジット、プロビット分析

- 順序ロジット(プロビット)モデルの定式化

(選択肢が3つの場合)



上記の場合、個人1 は選択肢 1 を、個人2は選択肢 3 を、個人3 と個人4 は選択肢 2 を選択する。

ロジット、プロビット分析

- 順序ロジット(プロビット)モデルの定式化(一般の場合)

今、$j = 1, \ldots, M$ のM 個の選択肢があり、これらの選択肢には 1, 2, 3, …, M と並べられるような何らかの自然な順序関係があるとする。

個人 i の特性を表す説明変数を X_i 、$U_i = \beta X_i + \varepsilon_i$ とすると

$$
\begin{aligned}
U_i < c_1 \quad &\Rightarrow \quad \text{選択肢 } 1 \text{ を選択} \\
c_{j-1} \le U_i < c_j \quad &\Rightarrow \quad \text{選択肢 } j \ (j = 2, \ldots, M-1) \text{ を選択} \\
c_{M-1} \le U_i \quad &\Rightarrow \quad \text{選択 } M \text{ を選択}
\end{aligned}
$$

となる。未知パラメーターは $c_1, c_2, \ldots, c_{M-1}, \beta$ である。

ロジット、プロビット分析

- 順序ロジット(プロビット)モデルの推定

最尤法で推定する。

Y_i を個人 i が 選択肢 j を選択したときに $Y_i = j$ を取る変数とする。ε_i の分布関数を $F_\varepsilon(.)$ とすると、個人 i が選択肢 1 を選択する確率は

$$\begin{aligned}\Pr(Y_i = 1) &= \Pr(\beta X_i + \varepsilon_i < c_1) \\ &= \Pr(\varepsilon_i < c_1 - \beta X_i) \\ &= F_\varepsilon(c_1 - \beta X_i)\end{aligned}$$

と表すことができる。

ロジット、プロビット分析

- 順序ロジット(プロビット)モデルの推定

同様に選択肢 j (j =2,…, M –1) が選択される確率は

$$
\begin{aligned}
\Pr(\,Y_i = j\,) &= \Pr(\,c_{j-1} \le \beta X_i + \varepsilon_i < c_j\,) \\
&= \Pr(\,c_{j-1} - \beta X_i \le \varepsilon_i < c_j - \beta X_i\,) \\
&= F_\varepsilon(c_j - \beta X_i\,) - F_\varepsilon(c_{j-1} - \beta X_i\,)
\end{aligned}
$$

と表せる。また、選択肢 Mが選択される確率は

$$
\begin{aligned}
\Pr(\,Y_i = M\,) &= \Pr\,(\,c_{M-1} < \beta X_i + \varepsilon_i) \\
&= \Pr(\,c_{M-1} - \beta X_i < \varepsilon_i\,) \\
&= 1 - F_\varepsilon\,(\,c_{M-1} - \beta X_i\,)
\end{aligned}
$$

となる。

ロジット、プロビット分析

- 順序ロジット(プロビット)モデルの推定

確率がわかれば、尤度関数および対数尤度関数は

(尤度関数)

$$L(\beta, c_1, \ldots, c_{M-1}) = \prod_{i=1}^{N} \prod_{j=1}^{M} \Pr(Y_i = j)^{1\{Y_i = j\}}$$

(対数尤度関数)

$$\log L(\beta, c_1, \ldots, c_{M-1}) = \sum_{i=1}^{N} \sum_{j=1}^{M} 1\{Y_i = j\} \log \Pr(Y_i = j)$$

で計算できる。対数尤度関数を最大化するようなパラメーター $\beta, c_1, \ldots, c_{M-1}$ が最尤推定値である。