

# 質的従属変数の ロジット、プロビット分析<sup>†</sup>

---

担当：長倉 大輔  
(ながくらだいすけ)

<sup>†</sup> 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

---

## ■ 多項ロジットモデル

先ほどは選択肢が車を「買う」、「買わない」のように2つの場合を考えた。実際の分析においては、選択肢は3つ以上ある場合が多い。

例えば夏休みに帰省する人々の交通手段を分析する場合などは、選択肢として「車」「新幹線」「飛行機」などが考えられ、さらに車の場合でも「高速道路を使う」か「一般道を使う」かに分かれ、「新幹線」の場合も「自由席」を使うか「グリーン車」を使うか、などと考えていくと、選択肢は非常に多くなる。

# ロジット、プロビット分析

---

## ■ 多項ロジットモデル

このように選択肢がいくつもあり、それぞれが選ばれる要因を分析する際には**多項ロジットモデル**と呼ばれるものがよく用いられる。これに対して先ほどのような選択肢が2つの場合のロジットモデルは**2項ロジットモデル**と呼ばれる。

多項ロジットモデルも2項ロジットモデルの時と考え方は同じであり、潜在変数モデルから導くことができる。

# ロジット、プロビット分析

---

## ■ 多項ロジットモデル

多項ロジットモデルは以下のような潜在変数モデルから導くことができる。

個人  $i$  は  $M$  個の選択肢の中から一つを選ぶとする。  
個人  $i$  が  $j$  ( $j = 1, \dots, M$ ) 番目の選択肢を選ぶことから得られる効用を  $U_{ij}$  とする。

これらの選択に影響を及ぼす変数を  $X_i$  とする  
(説明の簡単化のために1つだけとする)。

# ロジット、プロビット分析

---

## ■ 多項ロジットモデル

ここで効用  $U_{ij}$  は  $X_i$  の線形の関数、すなわち

$$U_{ij} = \alpha_j + \beta_j X_i + \varepsilon_{ij}, \quad (i = 1, \dots, N, j = 1, \dots, M)$$

と表せるとする。ここで  $\varepsilon_{ij}$  は  $X_i$  だけでは説明できない部分をまとめたものである。

ある選択肢  $k$  について、もし全ての  $j \neq k$  に対して

$$U_{ik} > U_{ij}$$

であるならば個人  $i$  は選択肢  $k$  を選ぶとする。

# ロジット、プロビット分析

---

## ■ 多項ロジットモデル

この時、個人  $i$  が選択肢  $k$  を選ぶ確率を  $F_{ik}$  とすると

$$\begin{aligned} F_{ik} &= \Pr ( U_{i1} < U_{ik}, U_{i2} < U_{ik}, \dots, U_{iM} < U_{ik} ) \\ &= \Pr ( \alpha_1 + \beta_1 X_i + \varepsilon_{i1} < \alpha_k + \beta_k X_i + \varepsilon_{ik}, \\ &\quad \alpha_2 + \beta_2 X_i + \varepsilon_{i2} < \alpha_k + \beta_k X_i + \varepsilon_{ik}, \dots \\ &\quad \dots, \alpha_M + \beta_M X_i + \varepsilon_{iM} < \alpha_k + \beta_k X_i + \varepsilon_{ik} ) \end{aligned}$$

となる。

# ロジット、プロビット分析

---

## ■ 多項ロジットモデル

2 項ロジットモデルの時と同様、 $\varepsilon_{i1}, \dots, \varepsilon_{iM}$  にある分布を仮定すると、この確率  $F_{ik}$  を計算する事ができる。

例えば、これらに独立な**タイプ1の極値分布**というものを仮定すると、 $F_{ik}$  は

$$F_{ik} = \frac{\exp(\alpha_k + \beta_k X_i)}{\sum_{j=1}^M \exp(\alpha_j + \beta_j X_i)}$$

となる。

# ロジット、プロビット分析

## ■ 多項ロジットモデルの尤度関数

観測されるデータは、個人  $i$  が  $j$  番目の選択肢をとれば  $Y_i = j$  となる変数を定義し

$$\{ Y_1 = 1, Y_2 = 3, Y_3 = 2, \dots, Y_N = M - 1 \}$$

等のように得られる。先ほどの  $F_{ik}$  を用いると

$$\begin{aligned} \Pr(Y_i = j) &= F_{i1}^{1\{Y_i=1\}} F_{i2}^{1\{Y_i=2\}} \dots F_{iM}^{1\{Y_i=M\}} \\ &= \prod_{j=1}^M F_{ij}^{1\{Y_i=j\}} \end{aligned}$$

と書く事ができる。ここで  $1\{Y_i=j\}$  とは  $Y_i=j$  ならば 1 を  $Y_i \neq j$  ならば 0 をとる関数である(指示関数と呼ばれる)

# ロジット、プロビット分析

---

- 多項ロジットモデルの尤度関数

よって  $\theta = \{ \alpha_1, \dots, \alpha_M, \beta_1, \dots, \beta_M \}$  とすると、尤度関数は

$$L(\theta) = \prod_{i=1}^N \Pr(Y_i = j) = \prod_{i=1}^N \prod_{j=1}^M F_{ij}^{1\{Y_i=j\}}$$

対数尤度関数は

$$\log L(\theta) = \sum_{i=1}^N \sum_{j=1}^M 1\{Y_i = j\} \log F_{ij}$$

となる。これを  $\theta$  について最大化して最尤推定値を得る。

# ロジット、プロビット分析

## ■ 識別性の問題

実際には  $\alpha_1, \dots, \alpha_M, \beta_1, \dots, \beta_M$  を推定することはできない。これはある選択肢  $p$  について ( $p$  は  $1, \dots, M$  のどれでもよい)  $\exp(\alpha_p + \beta_p X_i)$  で  $F_{ik}$  の分子と分母を割ることによって

$$F_{ik} = \frac{\exp[\alpha_k - \alpha_p + (\beta_k - \beta_p)X_i]}{\sum_{j=1}^M \exp[\alpha_j - \alpha_p + (\beta_j - \beta_p)X_i]} = \begin{cases} \frac{\exp(\gamma_k + \delta_k X_i)}{1 + \sum_{j \neq p} \exp(\gamma_j + \delta_j X_i)} & \text{for } k \neq p \\ 1 & \\ \frac{1}{1 + \sum_{j \neq p} \exp(\gamma_j + \delta_j X_i)} & \text{for } k = p \end{cases}$$

と表現できる事からわかる。ここで  $\gamma_j = \alpha_j - \alpha_p, \delta_j = \beta_j - \beta_p, \sum_{j \neq p}$  は  $p$  でない全ての  $j$  について和をとる事を意味する

# ロジット、プロビット分析

---

## ■ 識別性の問題

つまり  $F_{ik}$  は実際には  $\alpha_j$  と  $\alpha_p$  および  $\beta_j$  と  $\beta_p$  の「差」のみに依存しており、もともとの  $\alpha_j$  や  $\beta_j$  の大きさには依存していない。

よって実際に推定できるパラメーターはこの「差」である  $\gamma_j$  と  $\delta_j$  である。この時、対数尤度関数は前頁の  $F_{ik}$  をもとに構成する。

# ロジット、プロビット分析

---

- 説明変数が選択肢にも依存する場合

ここまで、潜在変数モデルにおいて効用 $U_{ij}$ は

$$U_{ij} = \alpha_j + \beta_j X_i + \varepsilon_{ij}, \quad (i = 1, \dots, N, \quad j = 1, \dots, M)$$

のように表せるとしてきた。

ここで $X_i$ は個人 $i$ の特性を表すもので、選択肢には依存しないと仮定してきた。

しかし実際には効用を説明する説明変数が個人 $i$ だけでなく**選択肢 $j$ にも依存する**ような場合が考えられる。

# ロジット、プロビット分析

---

## ■ (例) 通学手段の選択

例えば学生  $i$  がどの通学手段を選ぶかという問題を分析したいとしよう。

選択肢として、バス、電車、徒歩、バイク、の4つがあるとしよう。学生  $i$  の効用  $U_{ij}$  の説明変数として通学時間が考えられる。

通学時間は学生  $i$  がどこに住んでいるか、および、どの通学手段  $j$  を選ぶかの両方に依存すると考えるのが妥当であろう。

# ロジット、プロビット分析

---

## ■ (例) 通学手段の選択

すなわち、この時学生  $i$  の効用は

$$U_{ij} = \alpha_j + \beta_j X_{ij} + \varepsilon_{ij}$$

のように表す事ができる。ここで  $X_{ij}$  は学生  $i$  が通学手段  $j$  を選んだ時の通学時間である。

このような問題を分析する場合は**説明変数は個人と選択肢の両方に依存する**事になる。

# ロジット、プロビット分析

---

- 係数が共通の場合

ここで係数は選択肢  $j$  によって異なるとしているが、ある特定の問題においてはこの係数は選択肢に依存しないと考えることが自然な場合がある。

このような場合はさらに

$$U_{ij} = \alpha_j + \beta X_{ij} + \varepsilon_{ij}$$

と表す事ができる。すなわち  $X_{ij}$  の係数  $\beta$  はすべての選択肢について共通となっている。

# ロジット、プロビット分析

---

- 最も一般的な  $U_{ij}$  の形式

これらの場合を全て含むような最も一般的な  $U_{ij}$  の形式は

$$U_{ij} = \alpha_j + \beta X_{ij} + \gamma_j Z_i + \delta_j W_{ij} + \varepsilon_{ij}.$$

のように書く事ができる。

簡単化のために説明変数はそれぞれの場合に一つしかないとしているが、これらの変数がたくさんある場合に拡張するのは容易である。

# ロジット、プロビット分析

---

## ■ 識別性の問題 2

効用がこのように表せるとした時に、**識別できる**パラメータはどれであろうか？

選択は効用の差のみに依存しているのでこれがどのように表せるかを考えるとどのパラメータが識別可能かわかる。

# ロジット、プロビット分析

## ■ 識別性の問題 2

選択肢  $j$  の効用と選択肢  $k$  の効用の差は

$$\begin{aligned}U_{ij} - U_{ik} &= \alpha_j + \beta X_{ij} + \gamma_j Z_i + \delta_j W_{ij} + \varepsilon_{ij} \\ &\quad - (\alpha_k + \beta X_{ik} + \gamma_k Z_i + \delta_k W_{ik} + \varepsilon_{ik}) \\ &= \alpha_j - \alpha_k + \beta (X_{ij} - X_{ik}) + (\gamma_j - \gamma_k) Z_i \\ &\quad + \delta_j W_{ij} - \delta_k W_{ik} + \varepsilon_{ij} - \varepsilon_{ik} \\ &= \kappa_{jk} + \beta X_{ijk}^* + \omega_{jk} Z_i + \delta_j W_{ij} - \delta_k W_{ik} + \eta_{ijk}\end{aligned}$$

と書くことできる。ここで  $\kappa_{jk} = \alpha_j - \alpha_k$ ,  $X_{ijk}^* = X_{ij} - X_{ik}$ ,  $\omega_{jk} = \gamma_j - \gamma_k$ ,  $\eta_{ijk} = \varepsilon_{ij} - \varepsilon_{ik}$  である。

# ロジット、プロビット分析

## ■ 識別性の問題 2

これより識別できるパラメーターは(ある  $k$  に対して)  $\kappa_{jk}$ ,  $\beta$ ,  $\omega_{jk}$ ,  $\delta_j$ ,  $\delta_k$  であり、 $F_{ik}$  は。

$$F_{ik} = \begin{cases} \frac{\exp(\kappa_{kp} + \beta X_{ikp}^* + \omega_{kp} Z_i + \delta_k W_{ik} - \delta_p W_{ip})}{1 + \sum_{j \neq p} \exp(\kappa_{jp} + \beta X_{ijp}^* + \omega_{jp} Z_i + \delta_j W_{ij} - \delta_p W_{ip})}, & \text{for } k \neq p \\ \frac{1}{1 + \sum_{j \neq p} \exp(\kappa_{jp} + \beta X_{ijp}^* + \omega_{jp} Z_i + \delta_j W_{ij} - \delta_p W_{ip})}, & \text{for } k = p \end{cases}$$

となる。