離散選択モデル[1]

1　離散選択モデル

被説明変数が幾つかの限られた値を取るような状況では**離散選択（離散反応）モデル**（discrete choice, discrete response）が用いられる．もちろん厳密に言えば，所得や消費のようなデータでも整数の値しか取らないから離散的ではあるが，通常は多くても 10 程度の選択肢からひとつが選ばれるような状況を考える．離散選択モデルのなかでも，選ばれた値自身にはとくに意味のないモデルのことを**質的変数モデル**（qualitative variable）ともいう．離散選択モデルと質的変数モデルは同じものを指すようであるが，離散選択モデルには「回数」等の値に意味があるケースも含まれる．被説明変数が「Yes」「No」の 2 種類の値しか取らないケースは離散選択モデルの典型であり，とりうる値が 2 個の被説明変数をとくに 2 値変数（binary variable）とも呼ぶ．

離散選択モデルにもさまざまなモデルや推定法があるが，ここではそれぞれの離散選択が行われる確率を考え，その確率を最大にするようなパラメタを最尤法で推定するモデルに限定する．尤度あるいは条件付き尤度をそれぞれのモデルのもとで設定すれば，あとは最尤法によって推定が行われる．例として 2 値変数モデルを考えよう．被説明変数を y_i，説明変数ベクトルを x_i とする．y_i は 0 か 1 の値のみをとるとしよう．このとき，y_i が 1 となる確率を考え，

$$p(x_i) \equiv P(y_i = 1|x_i) \tag{15.1}$$

と書く．もちろん，y_i が 0 となる確率

$$P(y_i = 0|x_i) = 1 - P(y_i = 1|x_i) = 1 - p(x_i)$$

と書くことができる．ここで，確率分布関数 $P(.)$ の関数形を特定化できれば尤度を構成することができる．

離散選択モデルでしばしば興味の対象となるのは，説明変数 x_i の値の変化が被説明変数 y_i がある値を取る確率をどれほど変化させるかである．被説明変数自体は離散的な値しか取らないし，質的変数モデルではその数値自体には意味がないから，期待値の解釈は難しいことに注意しよう．2 値変数モデルのばあい，説明変数の確率への変化は

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ji}} = \frac{\partial p(x_i)}{\partial x_{ji}} \tag{15.2}$$

と表現できる．この値のことを**限界効果**（marginal effect）と呼ぶ．2 値変数モデルにおいてもこの限界効果を直接推定できることはあまりなく[2]，推定されたパラメタの値から一定

[1] Wooldridge (2002) *Econometric Analysis of Cross Section and Panel Data*, Ch.15 も参照せよ．ここで取り扱う内容は (15.1, 15.2, 15.3, 15.4, 15.5, 15.6, 15.9) の簡単なところのみである．式番号は Wooldridge (2002) による．

[2] 2 値変数モデルのうち，**線形確率モデル**（linear probability model）では，推定された係数がほぼ限界効果に対応する．ただし，線形確率モデルは，当てはめ値が 0 と 1 のあいだにおさまらなくなる可能性があることなどから，あまり用いられていないようである．

の仮定をおいて計算することが多い．逆に言えば，モデルはしばしば非線形であるために，推定されたパラメタの値そのものが解釈しやすい意味を持つことはあまりなく，その符号，あるいは一定の仮定のもとでさらに計算された値を解釈することになる．

離散選択モデルの推定にあたっては，連続変数としての**潜在変数**（latent variable）を想定すると便利なことがある．すなわち，潜在変数がある**範囲**の値を取れば質的変数がある値を取ると考える．また，潜在変数と説明変数の関係には，最も簡単なケースでは線形の関係を仮定し，その係数を推定する．係数の推定には，Wald，LR，LM 検定が用いられる．

ここでは，離散選択モデルのうちでも基本的な probit モデルと logit モデルを取り扱う．まず probit モデルを説明したのち，その応用としての順序 probit モデルを扱う．次に logit モデルと，その応用としての多項 logit，nested logit モデルを扱う．最後に，潜在変数の応用として区間回帰モデルにふれる．順序モデルや多項モデルについては，順序 logit や多項 probit も考えられるが，説明の簡単さのためにそれらは省略する．

2 Probit モデル

Probit モデルでは，確率分布関数 $P(.)$ として正規分布を用いる．平均ゼロ，分散 1 の標準正規分布の分布関数を $\Phi(.)$，確率密度関数を $\phi(.)$ で表す．すなわち，

$$\Phi(z) = \int_{-\infty}^{z} \phi(v)dv \tag{15.10}$$

$$\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2) \tag{15.11}$$

であり，この分布関数・密度関数について

$$\phi(z) = \Phi'(z), \quad \phi(z) = \phi(-z), \quad \phi'(z) = -z\phi(z)$$

が成り立つことが知られている．第 1 式は密度関数の定義，第 2 式は正規分布の対称性による．第 3 式は密度関数を微分すれば容易に求まる．

単純な probit モデル

単純な probit モデルとは，ここでは被説明変数が 2 値変数である 2 項選択モデルをいう．観測される変数 y_i は 0 か 1 の値しか取らない．ここで，0 と 1 という数値自体には意味がない．-1 と 3 の 2 種類の値しかとらない，と書いても議論はほとんど変わらないが，単に分かりにくくなるだけだろう．さて，観測される変数 y_i に対応して，観測されない潜在変数 y_i^* を考える．潜在変数 y_i^* は観測される変数 y_i が 1 をとるとりやすさの指標であり，潜在変数がある範囲の値を取れば質的変数が 1 となる，という関係にあるとする．ここでは，潜在変数が正の時には質的変数が 1，負の時には 0 の値を取るとしよう．すなわち，

$$\begin{aligned} y_i = 1 \quad & y^* > 0 \text{ のとき} \\ y_i = 0 \quad & y^* \leq 0 \text{ のとき} \end{aligned} \tag{15.9}$$

と書ける．もちろん，観測される変数の値が 0 から 1 へ変わる値（**閾値** threshold）はゼロでなくてもよいが，これをゼロ以外に設定しても定数項以外には影響しない．すなわち，説明変数から定数項を除外して，閾値を推定するという方法も考えられるが，係数についての結果は同じになる．

潜在変数は任意の実数値を取りうるとし，説明変数の線形関数であるとする.「線形関数である」とは推定されるパラメタベクトル β に対して線形であればよく，2 乗項・交差項・対数項等が入ってもよい．説明変数ベクトルを x_i とすると，

$$y_i^* = x_i\beta + u_i$$

と表現できる．u_i が互いに独立で同一の正規分布に従う誤差項であり，

$$u_i|x_i \sim N(0, \sigma^2)$$

とする．最小 2 乗推定のときには誤差項の分布の形状は仮定されていなかったことに注意しよう．

この設定のもとで，観測される変数が 1 あるいは 0 となる確率を求めよう．それぞれの観測値について質的変数が観測される確率（＝尤度）を求めることができれば，最尤法によって係数の推定を行うことができる．まず，観測される変数が 1 である確率は，

$$P(y_i = 1|x_i) = P(y^* > 0|x_i) = P(x_i\beta + u_i > 0|x_i) = P(u_i > -x_i\beta|x_i)$$

ここで，u_i は正規分布 $N(0, \sigma^2)$ に従うから，正規分布の対称性より

$$P(y_i = 1|x_i) = 1 - P(u_i < -x_i\beta|x_i) = 1 - (1 - P(u_i < x_i\beta|x_i)) = P(u_i < x_i\beta|x_i)$$

標準偏差で割って基準化すると，

$$P(y_i = 1|x_i) = P(u_i < x_i\beta|x_i) = \Phi\left(\frac{x_i\beta}{\sigma}\right)$$

同様に，観測される変数が 0 である確率は，

$$P(y_i = 0|x_i) = P(y^* \leq 0|x_i) = P(x_i\beta + u_i \leq 0|x_i) = P(u_i \leq -x_i\beta|x_i)$$

さきほどと同様の展開によって，

$$P(y_i = 0|x_i) = P(u_i \leq -x_i\beta|x_i) = 1 - \Phi\left(\frac{x_i\beta}{\sigma}\right)$$

u_i の独立性が仮定されれば，全体の尤度は各観測値の尤度の積で表現することができる．個々の条件付き尤度は

$$L_i = \left[\Phi\left(x_i\beta/\sigma\right)\right]^{y_i^*}\left[1 - \Phi\left(x_i\beta/\sigma\right)\right]^{1-y_i^*}$$

と表現できる．したがって個々の条件付き対数尤度は

$$\log L_i = y_i^* \log\left[\Phi\left(x_i\beta/\sigma\right)\right] + (1 - y_i^*)\log\left[1 - \Phi\left(x_i\beta/\sigma\right)\right]$$

ここで，関数形から分かるとおり，対数尤度を最大化する係数推定値 $\hat{\beta}$ の値は誤差項の分散 σ に依存するが，σ の値自体は決まらない．そこで，$\sigma = 1$ と置いて最大化を行う．これは，probit モデルでは潜在変数が閾値より大きいか小さいかのみが問題になっており，潜在変数の「大きさ」は問題にならないことに対応している．

観測値間の独立性が仮定されれば，対数尤度は個々の条件付き対数尤度の和となり，これを最大化して推定を行う．推定されるべきパラメタは係数ベクトル β である．推定されたベクトル $\hat{\beta}$ の値はどのような意味を持っているのだろうか．Probit モデルでは,

$$P(y_i = 1|x_i) = \Phi\left(x_i\hat{\beta}\right)$$

であるから，$\hat{\beta}$ の符号は潜在変数の変化の方向を示すとしても，値はそのままでは分かりにくい．限界効果は，しばしばデータの平均値 $\bar{x}$ で評価され,

$$\frac{\partial}{\partial x_{ki}} P(y_i = 1|\bar{x}) = \phi\left(\bar{x}\hat{\beta}\right)\hat{\beta}_k \tag{15.13}$$

も報告される．限界効果は確率の変化分であるので，その大きさは**%ポイント**で表す．たとえば，説明変数に平均値を代入したときの確率 $P(y_i = 1)$ が 70 %であるときにある説明変数の限界効果が 5 %ポイントである，とは，説明変数の値が 1 増加したときに，確率 $P(y_i = 1)$ が 70 %から 75 %に変化する，ということである．説明変数がダミー変数であるばあいには，限界効果よりも確率の差

$$P(y_i = 1|x_{ki} = 1) - P(y_i = 1|x_{ki} = 0)$$

のほうがわかりやすいかもしれない．実際，Stata の `dprobit` や `mfx` では，ダミー変数の説明変数については確率の差が出力される．

順序 probit モデル

順序反応モデル（ordered response）とは，被説明変数に採用される変数がとりうる選択肢に明確な順序が存在するようなばあいに用いられる．たとえば，なにかの好みが被説明変数であるとき,「好き」「嫌い」の 2 択であれば 2 項選択であるが,「好き」「やや好き」「どちらでもない」「やや嫌い」「嫌い」であれば選択肢は 5 つであり，この 3 つには順序が存在する．債券の格付け等もこの例に当てはまるし，資産運用の方針が「国債中心」「国債と株式混合」「株式中心」というのも順序反応モデルの対象となりうる．

順序の決まった観測される変数 y_i を規定する連続な潜在変数 y_i^* を想定しよう．被説明変数のとりうる値を J とする．観測される被説明変数の取りうる値が「好き」「やや好き」「どちらでもない」「やや嫌い」「嫌い」だとすれば（$J = 5$），潜在変数 y_i^* は「好き度」を示す連続変数である．潜在変数がある値（閾値）よりも大きな値となれば観測される被説明変数は「好き」となる，というように，潜在変数の一定の範囲に観測される被説明変数の値が対応していると考えよう．単純な probit モデルと同様に，この潜在変数が説明変数 x_i の 1 次関数で表現でき,

$$y_i^* = x_i\beta + u_i \tag{15.87}$$

と書けるとしよう．単純な probit モデルと同じく,「1 次関数である」とは推定されるパラメタベクトル β に対して線形であればよく，2 乗項・交差項・対数項等が入ってもよい．u_i は誤差項であり，正規分布に従うと仮定する．順序プロビットにおいても誤差項の分散は識別できないので，u_i は標準正規分布に従う，すなわち

$$u_i|x_i \sim N(0,1) \tag{15.87}$$

としよう．被説明変数のとりうる値の数は J 個だから，潜在変数の範囲を J 個に区切ってそれぞれに被説明変数の値が対応していると考えよう．J 個に区切るから区切りの数は $J-1$ 個であり，その値を小さいほうから $\alpha_1, \alpha_2, ..., \alpha_{J-1}$ としよう．対応する被説明変数の値をここでは $1, 2, ..., J$ とすると，潜在変数と被説明変数の対応は

$$\begin{aligned}
&y_i = 1 \quad \text{if} \quad y_i^* \le \alpha_1 \\
&y_i = 2 \quad \text{if} \quad \alpha_1 < y_i^* \le \alpha_2 \\
&y_i = 3 \quad \text{if} \quad \alpha_2 < y_i^* \le \alpha_3 \\
&\vdots \\
&y_i = J \quad \text{if} \quad \alpha_{J-1} < y_i^*
\end{aligned} \tag{15.88}$$

となる．ここで，閾値の値 $\alpha_1, \alpha_2, ..., \alpha_{J-1}$ も未知であることに注意しよう．被説明変数が 1 という値を取る確率は，単純な probit と同様に

$$P(y_i = 1|x_i) = P(y^* \le \alpha_1|x_i) = P(u_i \le \alpha_1 - x_i\beta|x_i) = \Phi(\alpha_1 - x_i\beta)$$

である．$y_i = 2$ となる確率も同じように考えると，

$$\begin{aligned}
P(y_i = 2|x_i) &= P(\alpha_1 < y_i^* \le \alpha_2|x_i) \\
&= P(\alpha_1 < x_i\beta + u_i \le \alpha_2|x_i) \\
&= P(u_i \le \alpha_2 - x_i\beta|x_i) - P(u_i < \alpha_1 - x_i\beta|x_i) \\
&= \Phi(\alpha_2 - x_i\beta) - \Phi(\alpha_1 - x_i\beta)
\end{aligned}$$

となる．$y_i = J$ については,

$$\begin{aligned}
P(y_i = 2|x_i) &= P(\alpha_{J-1} < y_i^*|x_i) \\
&= P(\alpha_{J-1} < x_i\beta + u_i|x_i) \\
&= 1 - P(u_i < \alpha_{J-1} - x_i\beta|x_i) \\
&= 1 - \Phi(\alpha_{J-1} - x_i\beta)
\end{aligned}$$

である．容易に分かるように，それぞれの値を取る確率を全て足すと 1 になる．被説明変数がそれぞれの値を取る確率が表現できたので，条件付き尤度関数を構成することができる．表現の簡単化のために，指標関数（indicator function）を導入しよう．指標関数とは,

カッコの中の条件が満たされているときだけ 1 であり，満たされていないときにはゼロの値を取る関数であり，$\mathbf{1}(.)$ で表す．個々の条件付き尤度は

$$L_i = [P(y_i = 1|x_i)]^{\mathbf{1}(y_i=1)} \times [P(y_i = 2|x_i)]^{\mathbf{1}(y_i=2)} \times ... \times [P(y_i = J|x_i)]^{\mathbf{1}(y_i=J)}$$

だから，先に求めた確率を代入すると個々の条件付き対数尤度は

$$\begin{aligned}\log L_i = &\ \mathbf{1}(y_i = 1)\log[\Phi(\alpha_1 - x_i\beta)] + \mathbf{1}(y_i = 2)\log[\Phi(\alpha_2 - x_i\beta) - \Phi(\alpha_1 - x_i\beta)] \\ &+ ... + \mathbf{1}(y_i = J)\log[1 - \Phi(\alpha_{J-1} - x_i\beta)] \end{aligned} \tag{15.89}$$

観測値の独立性が仮定されれば，サンプル全体の条件付き対数尤度は $\log L_i$ の和であり，これを最尤推定することができる．ここで推定されるパラメタは，潜在変数の係数ベクトル β と閾値の値 $\alpha_1, \alpha_2, ..., \alpha_{J-1}$ であるが，説明変数ベクトル x_i に定数項が含まれているときには閾値の値のうちの 1 つが識別されない．潜在変数の値と閾値の相対関係だけが問題となるからである．Stata のばあい，説明変数ベクトルに定数項が含まれない代わりに閾値がすべて推定される．

推定されたベクトル $\hat{\beta}$ の解釈について考えよう．この係数ベクトルは潜在変数の値を決めるから，係数 β_k が正に推定されれば，説明変数 x_{ki} が大きくなれば潜在変数 y_i^* の当てはめ値が大きくなることを示しており，したがって被説明変数 y_i も「大きく」なる傾向があることを表している．それゆえ，限界効果を説明変数の平均で評価することにすれば，単純なプロビットモデルと同じく，

$$\begin{aligned}\frac{\partial}{\partial x_{ki}} P(y_i = 1|\bar{x}) &= -\phi\left(\alpha_1 - \bar{x}\hat{\beta}\right)\hat{\beta}_k \\ \frac{\partial}{\partial x_{ki}} P(y_i = J|\bar{x}) &= \phi\left(\alpha_{J-1} - \bar{x}\hat{\beta}\right)\hat{\beta}_k\end{aligned}$$

が成り立つし,「端」でない $y_i = j$ についても

$$\frac{\partial}{\partial x_{ki}} P(y_i = j|\bar{x}) = \left[\phi\left(\alpha_{j-1} - \bar{x}\hat{\beta}\right) - \phi\left(\alpha_j - \bar{x}\hat{\beta}\right)\right]\hat{\beta}_k$$

が成り立つ.「端」でない $y_i = j$ となる確率の変化には，x_i の変化によって「下から入ってくる」要因と「上へ出て行く」要因の両方が影響することに注意しよう．

区間回帰モデル

順序 probit モデルの応用として，区間データ（interval-coded data）の推定がある．区間データとは，所得階層のように，一定の範囲に入っていることだけが情報として入手可能なデータである．たとえば所得データのばあい，実際の所得 y_i^* とコード化された所得階

層 y_i のあいだには

$$
\begin{aligned}
&y_i = 1 \quad \text{if} \quad y_i^* \le a_1 \\
&y_i = 2 \quad \text{if} \quad a_1 < y_i^* \le a_2 \\
&y_i = 3 \quad \text{if} \quad a_2 < y_i^* \le a_3 \\
&\vdots \\
&y_i = J \quad \text{if} \quad a_{J-1} < y_i^*
\end{aligned}
$$

のような関係が成り立つ．この関係は順序選択モデルにおける潜在変数と被説明変数の関係に似ているが，区間データのばあいは閾値 $a_1, a_2, ..., a_{J-1}$ は既知の数値であり，y_i^* が明確な意味を持っているという点で異なる．さて，y_i^* と説明変数との関係を検討するため，線形関係

$$y_i^* = x_i\beta + u_i$$

を想定し，このパラメタ β を推定したいとしよう．このとき，誤差項 u_i が説明変数の条件付きで正規分布に従うと仮定すれば，順序 probit モデルと同様にして尤度関数を構成し，最尤推定を行うことができる．ただし，閾値 $a_1, a_2, ..., a_{J-1}$ は既知の数値なので推定の対象とならず，また，誤差項の分散は 1 に基準化できず，こちらは推定の対象となる．区間回帰モデルでは，観測できない y_i^* が明確な意味を持ち，閾値が観測可能だから，$\log(y_i^*)$ を被説明変数とする線形関係を想定した回帰分析も可能となる．この場合には閾値も対数変換する必要がある．

3 Logit モデル

Logit モデルでは，確率分布関数 $P(.)$ としてロジスティック分布を用いる．分布関数を $\Lambda(z)$ と書くと，

$$\Lambda(z) = \frac{\exp(z)}{1 + \exp(z)} \tag{15.12}$$

である．

単純な logit モデル

単純な logit モデルとは，ここでは 2 項選択モデルをいう．潜在変数を想定し，probit のときと同じく

$$y_i^* = x_i\beta + u_i$$

とすれば，誤差項 u_i が標準ロジスティック分布に従うときとも解釈できる．logit モデルでは，観測される変数が 1 である確率は,

$$P(y_i = 1|x_i) = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)}$$

と表現され，推定されるべきパラメタは係数ベクトル β である．ここから対数尤度をただちに導くことができる．

Logit モデルと probit モデルはともに 2 項選択モデルである．最尤推定は定式化が正しければ推定量が一致性を持つから，逆に言えば，定式化が正しくなければ推定量の意義は怪しいものとなる．それゆえ，厳密に言えばサンプル (y_i, x_i) のデータ生成過程（DGP: data generating process）が logit モデルであるものを probit モデルで推定したり，その逆を行ったりすれば，推定量には信頼が置けないことになる．しかしじっさいには，分布の裾を除けば，いずれのモデルで推定しても，平均値周りで推定された限界効果は似たような値となることが多いし，最大化された対数尤度の値も似たようなものとなることが多い．それゆえ，単純なモデルを考えるかぎり，いずれのモデルを選択するかは実際の応用においてはほとんど問題とならない．もちろん，推定される係数ベクトル β の値は，関数形が異なるので，似たような数値にはならない．ただし，分布の中ほどについては,

$$\hat{\beta}_{\text{logit}} \simeq 1.6\hat{\beta}_{\text{probit}}$$

が成り立つことが知られている[3].

多項 logit モデル

離散選択モデルのうち，被説明変数がとりうる値が 3 つ以上あるときに，一般に多項選択モデル（multinomial）と呼ぶ．ここではそのうち，とりうる選択肢に明確な順序が存在しない（unordered response）ばあいを考える．職業選択や交通手段選択，学校選択等，その例は数多い．一見すると順序があるように見える労働供給量の選択にも多項選択モデルは応用されている．

被説明変数 y_i がとりうる選択肢が J 個あるとし，その属性が説明変数ベクトル x_i で表現されるような主体 i が J 個の選択肢から 1 つを選ぶという状況を考える．ここでは，それぞれの選択肢の属性が主体の選択に与える効果は捨象している．また，説明変数ベクトル x_i には定数項が含まれている．このとき，多項 logit モデルではそれぞれの選択肢 j を選ぶ確率は

$$P(y_i = j|x_i) = \frac{\exp(x_i\beta_j)}{1+\sum_{h=1}^{J-1}\exp(x_i\beta_h)} \quad \text{for } j = 1, ..., J-1 \tag{15.76}$$

と表される．経済主体は J 個の選択肢の中から 1 つ選んでいるから，J 個の選択確率の和は 1 に等しく，$j = J$ については

$$P(y_i = J|x_i) = \frac{1}{1+\sum_{h=1}^{J-1}\exp(x_i\beta_h)}$$

[3] Cameron, A. Colin, Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press., の (14.13) 式（p.473）による．

と基準化される．順序 probit モデルと同じく，指標関数 $\mathbf{1}(.)$ を用いればここから尤度関数を構成することができ，最尤推定を適用することができる．ここで推定されるパラメタはそれぞれの選択肢についての係数ベクトルたち $(\beta_1, ..., \beta_{J-1})$ であり，選択肢が 3 つあれば，2 つのベクトルが推定される．

推定されたパラメタ $(\beta_1, ..., \beta_{J-1})$ の解釈はなかなかめんどうである．説明変数 x_{ki} が連続変数であるとき，x_{ki} が 1 単位大きくなったときに選択肢 j が選ばれる確率の増分は

$$\frac{\partial}{\partial x_{ki}} P(y_i = j|x_i) = P(y_i = j|x_i) \left[\beta_{jk} - \frac{\sum_{h=1}^{J-1} \beta_{hk} \exp(x_i \beta_h)}{1 + \sum_{h=1}^{J-1} \exp(x_i \beta_h)} \right] \tag{15.77}$$

と計算される．もっと単純な解釈としては，相対的な選ばれ方（**オッズ比: odds ratio**）について

$$\frac{P(y_i = j|x_i)}{P(y_i = J|x_i)} = \exp(x\beta_j) \quad \text{for } j = 1, ..., J-1 \tag{15.78}$$

が成り立つから，この比の変分は $\beta_{jk} \exp(x\beta_i)\Delta x_k$ で近似される．また，同じことだが，対数オッズ比は線形結合 $x_i\beta$ で表現される．いずれにしても，β_{jk} の符号が正であれば，対応する説明変数 x_{ki} の値が大きくなれば選択肢 j が選ばれる確率が高くなることが分かる．また，定式化から明らかなように

$$P(y_i = j \text{ or } y_i = h|x_i) = P(y_i = j|x_i) + P(y_i = h|x_i)$$

なので，

$$P(y_i = j|y_i = j \text{ or } y_i = h, x_i) = \frac{\exp(x_i[\beta_j - \beta_k])}{1 + \exp(x_i[\beta_j - \beta_k])}$$

が成り立つ．

確率的選択モデル

ここまで述べてきた多項 logit モデルでは，選択肢のもつ属性の影響は考慮されておらず，選ぶ側の属性によってどの選択肢を選ぶ確率が高くなるか，のみが検討の対象であった．しかししばしば問題になるのは，選択肢の側の属性の影響であろう．どのような属性をもつ選択肢が選ばれやすいのか，という問題である．

このような問題は,（加法的）**確率効用モデル**（additive random utility）を基礎にした**確率的選択モデル**（probabilistic choice）によって分析される．いま，主体 i が選択肢 j を選んだときに得られる効用 y_{ij}^* が主体と選択肢ごとに定義される説明変数 x_{ij} と誤差項の線形関数で表されるとしよう．すなわち，

$$y_{ij}^* = x_{ij}\beta + a_{ij} \quad \text{for } j = 1, ..., J \tag{15.79}$$

y_{ij}^* は効用水準を表す潜在変数である．x_{ij} は主体ごと・選択肢ごとに異なる．たとえば x_{ij} は，個人 i が交通手段 j を選んだときの所要時間や，個人 i が病院 j に通うための交通費，

等である．主体ごと・選択肢ごとに異なる x_{ij} が観察でき，係数ベクトル β と誤差項 a_{ij} が決まれば，主体 i が選択肢 j を選んだときに得られる効用水準 y^*_{ij} が決まり，主体は効用が最も高くなるような選択肢を選ぶだろう．すなわち，

$$y_i = \mathrm{argmax}(y^*_{i1}, y^*_{i2}, ..., y^*_{iJ})$$

である[4]．いま，誤差項 a_{ij} が独立に同一の**タイプ I の極値分布（type I extreme value distribution）**に従う[5]とすれば，選択肢 j が選ばれる確率は

$$P(y_i = j|x_i) = \frac{\exp(x_{ij}\beta)}{\sum_{h=1}^{J}\exp(x_{ih}\beta)} \tag{15.80}$$

となる（McFadden 1974）．この確率を用いるモデルは条件付き選択モデル（conditional logit）とも呼ばれる．この確率値を偏微分してみると，選択肢 j の k 番目の属性 x_{jk} が変化したときに選択肢 j が選ばれる確率の変分は

$$\frac{\partial}{\partial x_{jk}}P(y_i = j|x_i) = P(y_i = j|x_i)[1 - P(y_i = j|x_i)]\beta_k \tag{15.81}$$

と表される．

多項 logit モデルとの比較

多項 logit モデルと確率的選択モデルは似たようなモデルであり，確率的選択モデルは多項選択モデルの一種とみなすこともあるが，選択肢の属性を考慮しているかしていないかという点で異なる．

多項 logit モデルのばあいには選択肢の属性は説明変数として含まれていないから，選択肢の属性が選択確率に与える影響は分析できない．したがって，選択肢の属性が重要でないか，興味の対象でないか，あるいは単に利用可能でないときに用いられる．家計のデータを集めて職業選択を分析する，といったばあいがこれにあたる．

確率的選択モデルには選択肢の属性が説明変数として含まれるから，家計や企業が観測可能な選択肢の属性に基づいて選択を行うとき，その観測可能な選択肢がどのような影響を与えるかが分析の対象となる．それゆえ，家計の購買行動（商品選択）や，仮想質問法の一種である**コンジョイント分析**（conjoint）に用いられる[6]．コンジョイント分析のばあいは観測可能な選択肢の属性を調査者が制御するが，観察データの場合にはデータの利用可能性に注意しなければならない．確率選択モデルの説明変数 x_{ij} は主体 i にとっての選択肢 j の値だから,「選ばれなかった選択肢」についての x_{ij} の値が必要であるからである．たとえば,「患者は近い病院に行く」という問題を確率的選択モデルで分析しようと思えば，行かなかったが選択肢に入っている病院を特定し，行かなかったその病院までの距離の情報を入手する必要がある．

[4] argmax は，その引数を最大にするように選ばれたものを表す記号．max は最大化された値を表す．
[5] タイプ I の極値分布の密度関数は，$f(z) = \exp(-z)\exp(-\exp(-z))$ である．
[6] コンジョイント分析のばあいには，選択している主体の属性を個別効果とみなして，panel logit モデルを用いるのが一般的のようである．

近年の応用では，選択する主体の属性も説明変数に含んだより一般的なモデルを用いる．多項 logit モデルでの説明変数になるような主体属性を表す説明変数ベクトルを w_i とすれば，確率的効用は

$$y^*_{ij} = x_{ij}\beta + w_i\delta_j + a_{ij} \quad \text{for } j = 1, ..., J$$

と表され（ただし $\delta_J = 0$），これをもとに尤度が構成される．

IIA：他の選択肢からの独立性

確率選択モデルは主体の選択行動をモデル化するのに便利な定式化であるが，制約もある．その制約として最も強いといわれているのが**他の選択肢からの独立性**（IIA: independence from irrelevant alternatives）である．選択肢 j が選ばれる確率は

$$P(y_i = j|x_i) = \frac{\exp(x_{ij}\beta)}{\sum_{h=1}^{J}\exp(x_{ih}\beta)} \tag{15.80}$$

であったから，2 つの選択肢 j, h が選ばれる相対的な確率は

$$\frac{P(y_i = j|x_i)}{P(y_i = h|x_i)} = \frac{\exp(x_{ij}\beta)}{\exp(x_{ih}\beta)} = \exp[(x_{ij} - x_{ih})\beta] \tag{15.83}$$

となり，問題となっている 2 つの選択肢 j, h 以外の選択肢の選ばれやすさとは無関係である．

他の選択肢からの独立性は応用問題によっては深刻な問題となる．また，ある選択肢が利用可能でなくなったときの相対的な選択確率は変化しないので，政策分析にも制約となりうる．極端な例を考えてみよう（McFadden 1974）．交通手段の選択を考える．最初の選択肢は車と赤いバスの 2 つであり，それぞれの選択確率は 1/2 ずつであるとしよう．ここに 3 番目の選択肢として青いバスが加わり，青いバスと赤いバスの相対的な選択確率が等しいとすれば，IIA の仮定のもとでは，車・赤いバス・青いバスの選択確率はすべて 1/3 となる．

このような問題の解決法もいろいろ提案されている．ひとつは，確率的効用の誤差項が任意の相関を持つ J 次の多変量正規分布（multinomial normal）に従うと仮定する多項probit モデルである．多項 probit モデルは，単純な probit モデルからの素直な拡張のように見えるが，J 次の多変量正規分布に基づく尤度関数は，高次の積分を含むため複雑なものとなり，実際上の計算は非常に困難となる．通常の最尤推定法では，選択肢が 5 個以上の推定は実際上不可能であるとされる．

いまひとつの解決法は，選択肢の構造を階層化したモデル（hierarchical）であり，入れ子型選択モデル（nested logit）がその代表である．このモデルでは，最終的な選択肢はいずれかのグループに分けられ，選択を行う主体は，第 1 段階目ではまずグループを選び，次にそのグループに含まれる選択肢の中から 1 つを選ぶと仮定される．グループ内の選択肢同士については IIA の仮定が必要となるが，グループを超えた選択肢については IIA の仮定が緩められる．

厚生評価

確率的選択モデルでは，潜在変数として効用値を想定しているから，経済厚生の評価が可能となる．ある選択肢の属性が変化したときに，それに伴う選択の変化をも考慮した経済厚生の評価が理論的には可能であり，しばしば補償変分（compensated variation）が用いられる．すなわち，選択肢の属性が変化したのちに，変化前の最大化された効用水準を達成するために必要な所得額を計算すればよい．もちろん，選択肢の属性の変化とともに，所得額の変化は選択を変化させ，最大化された効用水準を変化させるので，その評価は必ずしも容易ではない．

4 推定の評価

離散選択モデルについても，推定された係数の値以外にいくつかの統計量を報告し，その推定の評価を行う必要がある．係数推定値についての検定は Wald・LM・LR 検定を用いることができるから，通常の最小 2 乗推定のばあいと同じく，各推定値の標準誤差，係数がゼロという帰無仮説に対する有意水準を報告するのが普通である．これらについては，Stata 等のパッケージアプリケーションでは自動的に出力される．また，最大化された対数尤度 L も報告されることが多い．

最小 2 乗推定の決定係数 R^2 に対応するものの 1 つとして，perfect correctly predicted が推定のよさの指標として報告される．これは，推定結果から各観測値についてそれぞれの選択肢を選ぶ確率を求め，その確率が最も高いものが実現したとした結果と，実際の被説明変数の値が一致しているものの比率である．

決定係数に類似したものはいろいろ提案されており，まとめて pseudo R^2 と呼ばれる．2 項選択モデルでは，McFadden（1974）が pseudo R^2 を提案している．定数項のみを説明変数として含むモデルを推定し，その最大化された対数尤度を L_0 とし，実際に最大化された対数尤度 L に対して，$1 - L/L_0$ を決定係数とするものである．この値はゼロと 1 の間に収まる．パッケージアプリケーションでは pseudo R^2 が自動的に出力されるので，確認しておく必要があろう．

5 「内生性」問題

最尤推定は，尤度関数が正しく特定化されているときに一致性を持つから，そうでなければ推定結果は怪しいものとなる．推定結果が怪しいものとなる要因は，おおむね最小 2 乗法のときとよく似ている．ここではありうる問題について考えよう．

Stock and Watson (2006) では，最小 2 乗推定が一致性を持たないような状況（内的妥当性がない状況）として，Omitted variables，Misspesification of the functional form，Errors in variables，Sample selection，Simultaneous causality が挙げられていた．いずれも誤差項と説明変数の相関をもたらし，一致性を失わせる要因となった．

ここで扱ったモデルでは，誤差項が独立に同一の分布に従うと仮定して最尤推定を行っており，条件付き尤度の導出では，

$$u_i|x_i \sim N(0,1)$$

という条件を用いている．この条件は $E(u_i|x_i) = 0$ を含意するから，やはり誤差項と説明変数間の相関が問題となる．それゆえ，離散選択モデルにおいても説明変数と（潜在変数にかかわる）誤差項の相関に注意する必要があるし，問題となりそうな状況は最小 2 乗推定のときとそれほど変わらない．最小 2 乗推定のときには，直交条件が満たされず内的妥当性がないばあいには，適切な操作変数を探してきて 2 段階最小 2 乗法を行うという解決方法があった．離散選択モデルにおいては，2 段階最小 2 乗法はそのままは適用できず，原因に応じていろいろな手法が提示されている．そのいくつかは，見落とされている原因を明示的に数式で表現し，最尤推定によって追加的なパラメタをも推定する方法を採る．

たとえば，省略変数があるばあい，その**無視された異質性**（neglected heterogeneity）を表す誤差項以外に追加し，異質性の分布を仮定してそのパラメタを推定する手法もある（Wooldridge 2002, ch 15.7.1）．説明変数が逆の因果性を持つばあいには，同時方程式体系を明示的に考慮して尤度関数を構成すれば解決できることもある（Wooldridge 2002, ch 15.7.2-3）．操作変数法を応用した推定方法も提案されているし，サンプルがパネル構造であれば，個別効果を考慮した推定方法もありうる．尤度関数があまりに複雑になるばあいには，シミュレーションをともなう推定（maximum simulated likelihood）も用いられる．

6 Stata code

Probit モデル・logit モデルは Stata では以下のようなコマンドラインで推定することができる．

```
probit 被説明変数 説明変数
logit 被説明変数 説明変数
```

限界効果を求めるばあい，probit モデルでは `probit` のところを `dprobit` とすれば求めることができる．また，順序 probit モデル，多項 logit モデルのばあいは．

```
oprobit 被説明変数 説明変数
mlogit 被説明変数 説明変数
```

となる．