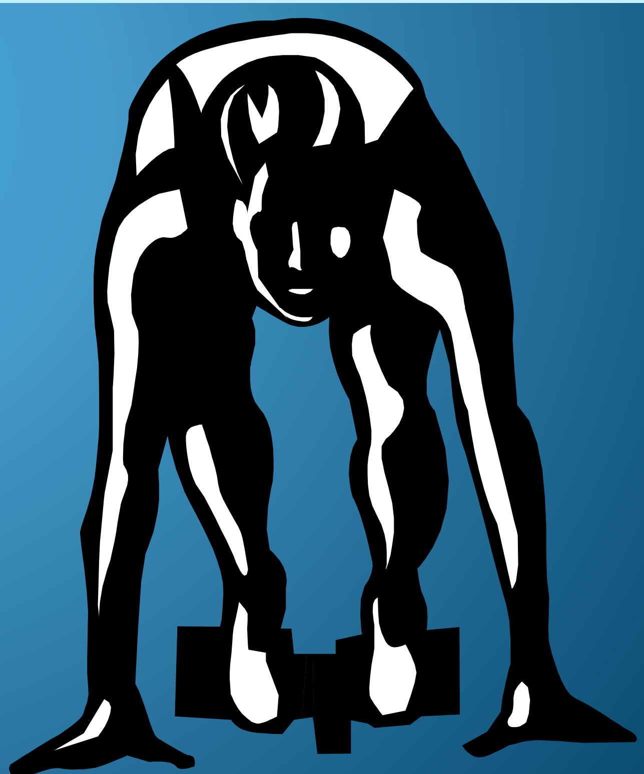
行動モデル編#1
MNLモデル･最尤推定法



2013/6/12　　理論勉強会#3

B4　森部伸一

MNLモデルについて

- MNLモデルとは、多項ロジットモデル（multinomial logit model）のことを指す

- 多項ロジットモデルの選択確率式が以下のようになることを、今回のゼミでは説明する

$$P_{n(i)} = \frac{\exp(\mu V_{in})}{\sum_{j=1}^{J} \exp(\mu V_{jn})} \quad , \mathrm{i} = 1, \ldots . \mathrm{J}$$

確率効用最大化モデル1

個人nの選択iに関する効用関数は以下のように設定できる

$$U_{in} = \beta_1 x_{1in} + \beta_2 x_{2in} + \cdots + \beta_K x_{Kin} + \varepsilon_{in}$$
$$= V_{in} + \varepsilon_{in}$$

U_{in}: 真の効用関数　V_{in}: 効用の確定部分　ε_{in}: 効用の確率項
x_{kin}: k番目の説明変数　β_k: k番目の未知パラメータ

効用の確率項に関しては以下の要因が含まれている
①確定項に含まれる変数以外の要因
②確定部分を線形和とした関数形の誤差
③属性の重みβ_kを個人間で均一とした誤差
④説明変数の測定誤差

確率効用最大化モデル2

ある個人がある選択肢を選択する

➡その個人にとってその選択肢の効用値が他の選択肢の効用値よりも大きい

個人nが選択肢iを選択する確率は以下のように表現できる

$$P_{n(i)} = \Pr\left[U_{in} \geq U_{jn}, for \forall j, i \neq j\right]$$

このように個人が効用を最大化するとして導かれるモデルを確率効用最大化モデル（random utility maximization(RUM)）と呼ぶ

ガンベル(Gumbel)分布１

ロジットモデルでは、効用の確立項εはガンベル分布を仮定

＜累積分布関数＞

$$F_{(\varepsilon)} = \exp[-\exp\{-\mu(\varepsilon - \eta)\}]$$

＜確率密度関数＞

$$f_{(\varepsilon)} = \mu \exp\{-\mu(\varepsilon - \eta)\} \exp[-\exp\{-\mu(\epsilon - \eta)\}]$$

μ：分布のばらつきを表すスケールパラメータ

η：分布の位置を表すロケーションパラメータ

ガンベル(Gumbel)分布２

ガンベル分布を次のような性質を持っている

①最頻値：η　**平均値：**$\eta+\gamma/\mu$(**γはオイラー定数**≅0.577)

②分散：$\frac{\pi^2}{6\mu^2}$

③ε_1**と**ε_2**がそれぞれ**(η_1, μ), (η_2, μ)**のパラメータを持つ互いに独立なガンベル分布に従うとき、**$\varepsilon \geq \varepsilon_1 - \varepsilon_2$**である確率は**

$$\Pr[\varepsilon_1 - \varepsilon_2 \leq \varepsilon] = \frac{1}{1 + \exp\{\mu(\eta_2 - \eta_1 - \varepsilon)\}}$$

ガンベル(Gumbel)分布３

④ $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_J$がそれぞれ$(\eta_1, \mu)$,$(\eta_2, \mu)$,…,$(\eta_J, \mu)$のパラメータをもつ互いに独立なガンベル分布に従うときは、$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_J$の最大値もガンベル分布に従い、そのパラメータは以下のようになる

$$\left(\frac{1}{\mu} ln \sum_{j=1}^{J} \exp(\mu\eta_j), \mu\right)$$

MNLモデル１

各選択肢の効用の確率項εがそれぞれパラメータ(o,μ)を持つ互いに独立なガンベル分布に従うとする。

個人nが選択肢1を選択する確率を考える

$$P_{n(1)} = \Pr\left[V_{1n} + \varepsilon_{1n} \geq max_{j=2,3,\dots,J}\left(V_{jn} + \varepsilon_{jn}\right)\right]$$

$U_{2n}, U_{3n}, \dots, U_{Jn}$がそれぞれ$(V_{2n}, \mu)$,$(V_{3n}, \mu)$,…,$(V_{Jn}, \mu)$のパラメータをもつ互いに独立なガンベル分布に従うので、

${U_n}^* = max_{j=2,3,\dots,J}\left(V_{jn} + \varepsilon_{jn}\right)$とすると性質④より${U_n}^*$は以下のパラメータに従うガンベル分布になる。

$$\left(\frac{1}{\mu} ln \sum_{j=2}^{J} \exp\left(\mu V_{jn}\right), \mu\right)$$

MNLモデル2

$U_n{}^* = V_n{}^* + \varepsilon_n{}^*$として、$V_n{}^* = \frac{1}{\mu} ln \sum_{j=2}^{J} \exp(\mu V_{jn})$とおくと

$\varepsilon_n{}^*$はパラメータ(o,μ)を持つガンベル分布に従う

$$P_{n(1)} = \Pr[V_{1n} + \varepsilon_{1n} \geq V_n{}^* + \varepsilon_n{}^*]$$
$$= \Pr[\varepsilon_n{}^* - \varepsilon_{1n} \leq V_{1n} - V_n{}^*]$$

ε_{1n}と$\varepsilon_n{}^*$がそれぞれ(o,μ), (o,μ)のパラメータを持つ互いに独立なガンベル分布に従うとき、$V_{1n} - V_n{}^* \geq \varepsilon_n{}^* - \varepsilon_{1n}$である確率は、性質③より

$$P_{n(1)} = \frac{1}{1+\exp\{\mu(0-0-(V_{1n}-V_n{}^*))\}}$$ となる

MNLモデル3

$$P_{n(1)}=\frac{1}{1+\exp\{\mu(0-0-(V_{1n}-V_n{}^*))\}}=\frac{1}{1+\exp\{\mu(V_n{}^*-V_n)\}}$$

$$=\frac{\exp(\mu V_{1n})}{\exp(\mu V_{1n})+\exp(\mu V_n{}^*)}$$

$$=\frac{\exp(\mu V_{1n})}{\exp(\mu V_{1n})+\exp\left(ln\sum_{j=2}^{J}\exp\left(\mu V_{jn}\right)\right)}\ (\because V_n{}^*=\frac{1}{\mu}ln\sum_{j=2}^{J}\exp\left(\mu V_{jn}\right))$$

$$=\frac{\exp(\mu V_{1n})}{\sum_{j=1}^{J}\exp\left(\mu V_{jn}\right)}$$

多項ロジットモデルの選択確率式

最尤推定法について1

離散選択モデルにおける未知パラメータ

効用関数の説明関数にかかる未知係数(先の例β)

誤差項のスケールパラメータ(先の例μ)

離散選択モデルにおける未知パラメータは、アンケート集計などで得られた推計データにもっとも一致するように定められるべき

パラメータ値を求める方法として、最尤推定法(maximum likelihood estimation; MLE)が最も一般的に用いられる

最尤推定法について２

最尤推定法の考え方

行動を表す理論モデル(ここに未知パラメータが含まれている)が正しいと仮定の下で、観測されたデータが得られるもっともらしさ(尤度(likelihood))が最大になるようにパラメータを定めることで、パラメータを推定する。

最尤推定法の例1

コインを続けて6回振ったところ、そのうち5回が表であった。
このコインの表の出る確率qはいくらか。

最尤法の考え方

確率qを変化させ、観測された事象が実現する確率を最大にすることで、確率qの値を推定する。

尤度関数(観測された現象が実現する確率)

$$L(q) = {}_6C_5 q^5(1-q) = 6q^5(1-q)$$

尤度関数が最大になればいい

最尤推定法の例2

尤度関数を最大にするときは？

$L(q) = 6q^5(1-q)$ は右のような形状なので、微分して傾きが0になるときを求めればよい。

$\frac{dL(p)}{dq}$=$6q^4(5-6q)$=0　より$q = \frac{5}{6}$

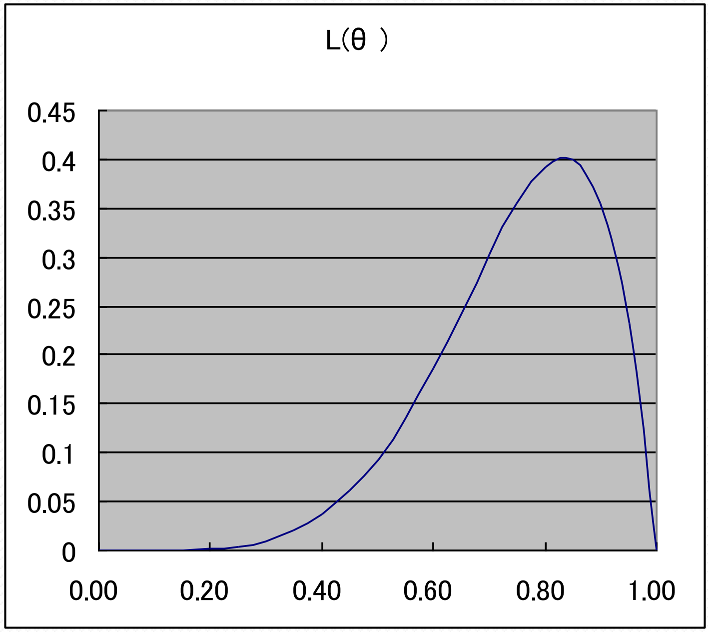


図1

あるいは対数を取って微分をしてもよい

$\ln L(q) = \ln 6 + 5\ln q + \ln(1-q)$

$\frac{d\ln L(q)}{dq} = \frac{5-6q}{q(1-q)} = 0$ より求めることが出来る

最尤推定法の例3

離散選択モデルのもとで、データの尤度(もっともらしさ)は選択確率を用いて以下のように表される

$$L = \prod_{n=1}^{N} \prod_{i=1}^{J} P_n(i)^{d_{in}}$$

d_{in}:個人nが選択肢iを選択したとき1, そうでないとき0

今回は、
二項のロジットモデル(選択肢1, 2)でパラメーター推定について考えてみる

この尤度関数は値が小さく、数値計算誤差が生じやすいため、尤度関数の対数を取ったものの最大化を考える。

$$\sum_{n=1}^{N} \sum_{j=1}^{2} d_{in} \ln P_n(i)$$

最尤推定法の例4

選択肢1、2(バス、自家用車)として

$V_{1n} = \alpha t_{1n} + \beta C_{1n}$

$V_{2n} = \alpha t_{2n} + \beta C_{2n} + \gamma$

と定める。

また、選択状況を右の表のように仮定する。

t1	t2	c1	c2	1	2
20	5	210	50	10	120
20	10	210	100	30	90
30	15	210	100	10	60
30	20	210	125	10	60
45	30	210	150	30	90
45	45	210	250	60	70
60	60	420	400	20	20
60	30	420	1100	25	10
60	45	420	1100	30	5

最尤推定法の例5

選択確率は以下のようになる

$$P_{1n} = \frac{\exp(\alpha t_{1n}+\beta C_{1n})}{\exp(\alpha t_{1n}+\beta C_{1n})+\exp(\alpha t_{2n}+\beta C_{2n}+\gamma)}$$

$$P_{2n} = 1 - P_{1n}$$

$$\sum_{n=1}^{N}\sum_{j=1}^{2} d_{in} \ln P_{n}(i)$$

d_{in}:個人nが選択肢iを選択したとき1, そうでないとき0

が最大になるようにパラメーターを求める。

最尤推定法の例6

Excelのソルバーという機能を用いて、最大となる解を用いると以下のようになる。

t1	t2	c1	c2	n1	n2	p1	p2	ln(p1)*n1	ln(p2)*n2
20	5	210	50	10	120	0.127198	0.872802	-20.6201	-16.3256
20	10	210	100	30	90	0.201584	0.798416	-48.0464	-20.2613
30	15	210	100	10	60	0.154616	0.845384	-18.6681	-10.0778
30	20	210	125	10	60	0.220481	0.779519	-15.1194	-14.9447
45	30	210	150	30	90	0.186679	0.813321	-50.3509	-18.5967
45	45	210	250	60	70	0.487448	0.512552	-43.1143	-46.7847
60	60	420	400	20	20	0.420004	0.579996	-17.3498	-10.8947
60	30	420	1100	25	10	0.715512	0.284488	-8.36893	-12.5706
60	45	420	1100	30	5	0.868709	0.131291	-4.22243	-10.1517
α	β	γ						-225.861	-160.608
-0.06449	-0.00454	0.231912							-386.468

まとめ

多項ロジットモデルは、ガンベル分布の性質によって確定項のみで選択確率が求められる。

最尤法は、もっともらしさ(実際の現象が起きるような確率)が最大になるようにパラメーターを定めることで、パラメーターの推定を行っている。

補足

ガンベル分布　性質③について

$\varepsilon_1, \varepsilon_2$がパラメータ$(\eta_1, \mu)$,$(\eta_2, \mu)$を持つ独立のガンベル分布に従うとき、$\varepsilon_1 - \varepsilon_2 \leq \varepsilon$の確率について考える。

$$\mathrm{F}(\varepsilon) = \mathrm{Prob}(\varepsilon_1 - \varepsilon_2 \leq \varepsilon) = \mathrm{Prob}(\varepsilon_1 \leq \varepsilon + \varepsilon_2)$$

$$= \int_{\varepsilon_2=-\infty}^{\infty} \int_{\varepsilon_1=-\infty}^{\varepsilon+\varepsilon_2} f\,(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2 = \int_{\varepsilon_2=-\infty}^{\infty} F\,(\varepsilon + \varepsilon_2) f(\varepsilon_2) d\varepsilon_2$$

$$= \int_{-\infty}^{\infty} e^{-e^{-\mu(\varepsilon+\varepsilon_2-\eta_1)}} \mu e^{-\mu(\varepsilon_2-\eta_2)}\, e^{-e^{-\mu(\varepsilon_2-\eta_2)}} d\varepsilon_2 = \int_{-\infty}^{\infty} \mu e^{-\mu(\varepsilon_2-\eta_2)}\, e^{-e^{-\mu\varepsilon_2}(e^{-\mu\varepsilon+\mu\eta_1}+e^{\mu\eta_2})} d\varepsilon_2$$

ここで$\delta = e^{-\mu\varepsilon+\mu\eta_1} + e^{\mu\eta_2}$とおく

$$= \int_{-\infty}^{\infty} \mu e^{-\mu(\varepsilon_2-\eta_2)} e^{-\delta e^{-\mu\varepsilon_2}}\ d\varepsilon_2 = \int_{-\infty}^{\infty} \frac{1}{\delta} e^{\mu\eta_2} \{\mu\delta e^{-\mu\varepsilon_2} e^{-\delta e^{-\mu\varepsilon_2}}\}\ d\varepsilon_2$$

$$= \frac{1}{\delta} e^{\mu\eta_2} \int_{-\infty}^{\infty} (e^{-\delta e^{-\mu\varepsilon_2}})'\ d\varepsilon_2 = \frac{1}{\delta} e^{\mu\eta_2} \left(e^0 - \frac{1}{e^\infty}\right) = \frac{1}{\delta} e^{\mu\eta_2}$$

$$= \frac{e^{\mu\eta_2}}{e^{-\mu\varepsilon+\mu\eta_1} + e^{\mu\eta_2}} = \frac{1}{1 + e^{\mu(\eta_1-\eta_2-\varepsilon)}}$$