

## 非集計行動モデルの集計誤差について メモ

### 1. 序

非集計モデルの集計誤差については、研究すべきテーマは 1980 年代初頭に完遂したと思われる。アメリカでは 1970 年代半ばに Koppelman(1976)により、そして日本では当時北海道大学の山形・桐越らにより精力的な研究が積み重ねられてきた。それ以降、「集計方法は分類法で十分」という結論が実務上でも一般的になり、集計誤差自体を推計することも行われなくなった。しかし今回、修士論文で航空の Revenue Management (RM)を扱う過程で、Logit モデルの集計誤差を定量的に計算する必要に迫られた。簡単に言えば、航空動態調査結果から「券種選択モデル」を構築し、クラス（券種）別の利用者を予測し、航空会社のクラス別最適価格と座席配分を求める典型的な RM アルゴリズム（EMSR）を考えていた。この RM ではクラス別需要と共にその分散値が支配的役割を果たす。そこで Logit モデルの集計誤差について理論に基づく推計が必要となった。

調べていくうちに、最近の文献では集計誤差の定義やその計算例について全く触れていないことが分かった。唯一、Ben-Akiva et al.(1985)のテキストで Koppelman(1976)を引用して、その基礎式を紹介しているが、数式の標記に誤りがあったり（兵藤が持っているのは 1985 年発行版）計算例が示されていないなど、集計誤差の重要性に比して十分な記述がなされていない。

大半の文献が 1980 年前後であったため、若干その理論を整理するのに時間を要したが、上記の理由から、ここに集計誤差推計方法について例を示しながらまとめておくことにする。

### 2. 誤差伝播の法則について

集計誤差を計算する上で、もっとも重要な基礎式が誤差伝播式である。土木工学科では測量学の授業で習うことが多いが、式が簡単な構造であることもあり、単なる丸暗記で覚えていることが多いのではないだろうか？（兵藤もそうだった）しかも測量学では、共分散項は省いて教えるようで、ますます式の導出過程が不明確になるようだ。まずは基礎式を整理しよう。

2 変数で表される関数、 $y = f(\theta_1, \theta_2)$  について考える。関数を点、 $(\theta_{10}, \theta_{20})$  の周りで一次のテーラー展開を行う。

$$y - y_0 = \frac{\partial f}{\partial \theta_1} (\theta_1 - \theta_{10}) + \frac{\partial f}{\partial \theta_2} (\theta_2 - \theta_{20}) \quad (1)$$

ここで、 $y - y_0$  の 2 乗値の期待値が  $y$  の分散になることから期待値式を導く。

$$E[(y - y_0)^2] = \left( \frac{\partial f}{\partial \theta_1} \right)^2 E[(\theta_1 - \theta_{10})^2] + \left( \frac{\partial f}{\partial \theta_2} \right)^2 E[(\theta_2 - \theta_{20})^2] + 2 \cdot \left( \frac{\partial f}{\partial \theta_1} \right) \left( \frac{\partial f}{\partial \theta_2} \right) \cdot E[\theta_1 - \theta_{10}] \cdot E[\theta_2 - \theta_{20}] \quad (2)$$

分散記号を用いて書き直すと、

$$\sigma_y^2 = \left(\frac{\partial f}{\partial \theta_1}\right)^2 \sigma_1^2 + \left(\frac{\partial f}{\partial \theta_2}\right)^2 \sigma_2^2 + 2 \cdot \left(\frac{\partial f}{\partial \theta_1}\right) \left(\frac{\partial f}{\partial \theta_2}\right) \sigma_{12}^2 \quad (3)$$

というおなじみの誤差伝播式を得ることができる。この式を一般的な多変量関数に拡張すれば、その一般式は下記の通りとなる。

$$\sigma_y^2 = \sum_i \sum_j \left(\frac{\partial f}{\partial \theta_i}\right) \left(\frac{\partial f}{\partial \theta_j}\right) \sigma_{ij}^2 \quad (4)$$

ここで  $\sigma_{ij}^2$  は母数パラメータ  $\theta_i$  と  $\theta_j$  間の分散共分散行列である。

### 3. Logit モデルの集計誤差式

集計方法として平均値法を考える。

$$P_i = \frac{e^{V_i}}{\sum_j e^{V_j}} \quad (5)$$

上記 Logit モデル式について効用項の微小変化を考え、選択確率に関する誤差伝播式を定義すれば、(4)式と(5)式から、

$$\text{var}[P_i] = \sum_j \sum_k \left(\frac{\partial P_i}{\partial V_j}\right) \left(\frac{\partial P_i}{\partial V_k}\right) \text{cov}[V_j, V_k] \quad (6)$$

を得る。ちなみに、

$$\frac{\partial P_i}{\partial V_j} = \begin{cases} P_i(1-P_i) & \text{if } i=j \\ -P_i P_j & \text{otherwise} \end{cases} \quad (7)$$

である。あとは効用項の分散共分散行列  $\text{cov}[V_j, V_k]$  を求めればよい。

効用項にはパラメータと説明変数が含まれ、それぞれが誤差分散を有する。パラメータの分散共分散行列は情報行列の逆行列として、パラメータ推定結果と共に得ることができる。また、説明変数の分散共分散行列は事後的に設定する必要がある。例えば OD 間の所要時間であればゾーニングに起因する誤差や、時間帯別の変動などに起因する誤差であり、車保有率などの社会経済属性変数であれば、標本調査による誤差（平均値分散式から算出される）として定義可能であろう。

さて、単回帰  $y=ax$  であれば、右辺の分散は  $a^2 \text{var}[x] + x^2 \text{var}[a]$  で定義されるが、同様に効用項の誤差を考える。まず以下の条件や変数を定義する。

選択肢の数：nm      説明変数の数：np

説明変数： $\mathbf{X}$  (np 行×nm 列)      パラメータベクトル： $\boldsymbol{\beta}$  (np 行×1 列)

選択肢  $i$  と選択肢  $j$  間の説明変数分散共分散行列： $\boldsymbol{\Sigma}_{ij}$  (np 行×np 列)

パラメータ分散共分散行列： $\boldsymbol{\Sigma}_\beta$  (np 行×np 列)

ここでは説明変数の分散共分散行列を後の式標記を単純化するため、選択肢間の組み合わせで分解標記することとする。

効用の誤差分散は単回帰の例と同様、パラメータ誤差に起因する項と、説明変数誤差に起因する項に分けられ、それぞれ選択肢間の共分散も考慮する必要がある。前者の誤差項からなる選択肢間分散共分散行列を  $\Lambda_1$  (nm 行×nm 列) とすれば、

$$\Lambda_1 = X \Sigma_{\beta} X \tag{8}$$

一方、後者の誤差項からなる選択肢間分散共分散行列の  $i$  行  $j$  列要素を  $\Lambda_{2ij}$  (スカラー) とすれば、

$$\Lambda_{2ij} = \beta' \Sigma_{x_{ij}} \beta \tag{9}$$

となる。ここで、(9)式を要素とする選択肢間の（説明変数誤差に起因する）分散共分散行列を  $\Lambda_2$  (nm 行×nm 列) と再定義する。すると、効用項の誤差分散行列は(8)式、(9)式から、 $\Lambda_1 + \Lambda_2$  となる。これと、(6)式より、各選択肢の誤差分散を推計することができる。

#### 4. 計算事例

代表的な都市間交通手段選択データを用いて、集計誤差の推計を試みる。用いたのは 1995 年の旅客純流動調査からランダムに抽出した長距離帯の 6,000 サンプル（仕事目的 3,000、その他目的 3,000 サンプル）で、対象交通手段は「航空・鉄道・車」の 3 手段である。

MNL モデルの推定結果を以下にまとめる。

表-1 パラメータ推定結果

	Para.	STD.	t値		
時間 (共通)	-0.9270	0.0252	-36.8	サンプル数	6000
費用 (共通)	-3.6276	0.6377	-5.7	初期尤度	-6005.0
目的D (Air)	0.5230	0.0957	5.5	最終尤度	-4475.7
目的D (Rail)	0.0180	0.0882	0.2	尤度比	0.2543
定数項 (Air)	-2.1877	0.1029	-21.3	的中率	67.2%
定数項 (Rail)	-0.7166	0.0717	-10.0		

(時間は 100 分、費用は 10 万円単位。目的は「仕事=1」)

表-2 説明変数の選択肢別平均値・標準偏差

		Ave.	Std.
Air	時間	1.842	0.468
Air	費用	0.230	0.084
Air	目的D	0.500	0.500
Air	0	0	0
Air	定数項	1	0
Air	0	0	0
Rail	時間	4.097	3.942
Rail	費用	0.173	0.067
Rail	0	0	0
Rail	目的D	0.500	0.500
Rail	0	0	0
Rail	定数項	1	0
Car	時間	6.871	4.422
Car	費用	0.269	0.212
Car	0	0	0
Car	0	0	0
Car	0	0	0
Air	share	0.440	
Rail	share	0.420	
Car	share	0.140	

		時間 (共通)	費用 (共通)	目的D (Air)	目的D (Rail)	定数項 (Air)	定数項 (Rail)
時間	(共通)	6.35E-04	-4.11E-04	-2.22E-04	4.79E-05	1.79E-03	8.32E-04
費用	(共通)	-4.11E-04	4.07E-01	2.88E-04	-2.95E-03	-1.79E-02	6.55E-03
目的D	(Air)	-2.22E-04	2.88E-04	9.15E-03	6.25E-03	-5.48E-03	-3.51E-03
目的D	(Rail)	4.79E-05	-2.95E-03	6.25E-03	7.79E-03	-2.96E-03	-3.91E-03
定数項	(Air)	1.79E-03	-1.79E-02	-5.48E-03	-2.96E-03	1.06E-02	5.27E-03
定数項	(Rail)	8.32E-04	6.55E-03	-3.51E-03	-3.91E-03	5.27E-03	5.14E-03

表-3 推定パラメータの分散共分散行列

表-4 説明変数間の相関係数行列

	Air 時間	Air 費用	Air 目的D	Air 0	Air 定数項	Air 0	Rail 時間	Rail 費用	Rail 0	Rail 目的D	Rail 0	Rail 定数項	Car 時間	Car 費用	Car 0	Car 0	Car 0	Car 0
Air	時間	1.000																
Air	費用	0.750	1.000															
Air	目的D	0.000	0.000	1.000														
Air	0	0.000	0.000	0.000	1.000													
Air	定数項	0.000	0.000	0.000	0.000	1.000												
Air	0	0.000	0.000	0.000	0.000	0.000	1.000											
Rail	時間	0.432	0.623	0.000	0.000	0.000	0.000	1.000										
Rail	費用	0.592	0.861	0.000	0.000	0.000	0.000	0.760	1.000									
Rail	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000								
Rail	目的D	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000							
Rail	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000						
Rail	定数項	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000					
Car	時間	0.553	0.824	0.000	0.000	0.000	0.000	0.899	0.923	0.000	0.000	0.000	0.000	1.000				
Car	費用	0.473	0.758	0.000	0.000	0.000	0.000	0.925	0.883	0.000	0.000	0.000	0.000	0.973	1.000			
Car	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000		
Car	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	
Car	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Car	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

さて、説明変数の分散共分散行列の設定方法には幾つかの議論を必要とする。表-4 はパラメータ推定に用いた 6,000 サンプルのデータから得られた相関係数行列の結果である。見いだされるポイントとしては、

- ・ 時間、費用間の相関は同じ選択肢間のみならず、異なる選択肢間でも大きい
- ・ 選択肢間で同じ値を持つ社会経済変数は、選択肢間の相関係数は 1 となる
- ・ 定数項の共分散は 0 である

などがあげられる。さて、集計予測を行う場合、説明変数の分散共分散をどのように設定すべきか？ 代表的な説明変数に応じて考察してみる。

a)所要時間：ゾーニングに起因する誤差（アクセス、イグレス時間のバラツキが大きい）が支配的で

あると考えられる。その他、自動車であれば時間帯別の変動も大きな要素であろう。

b)費用：所要時間ほど大きな誤差はないように思われるが、実際には航空運賃の割引構成率が不明であるなど、主に割り引き運賃設定に関わる誤差が大きい。都市圏内では自己負担、会社負担別のいわば実質費用に関わる誤差要因がある。

c)社会経済変数：年齢・性別・車保有など、ゾーン単位の正確なデータが存在する場合は、予測対象 OD 交通量と母集団（発生ゾーンの人口など）との関係で誤差が生じる。すなわち、OD 交通量が母集団からの標本抽出であると考えた場合、いわゆる比率の平均値分散式を用いた

$$\sigma_p^2 = \frac{N-n}{N-1} \cdot \frac{p(1-p)}{n} \quad (10)$$

を説明変数の分散とすることが考えられる。

本稿では、集計誤差式の提示を主目的とするため、上記説明変数の誤差について詳細な議論に立ち入らない。まずもって、表-4 の相関係数行列は、距離帯分布が大きい都市間交通の OD ペア全体の相関であり、予測時に発生する単一 OD ペアの誤差とは発生要因が全く異なることに留意すべきである。

これら問題点を認識した上で、説明変数の分散共分散行列として、下記の設定を行う。

- ・ 説明変数の平均値は 6,000 サンプルの平均値（表-2）を用いる
- ・ 所要時間、費用に関しては変動係数(C.V.)を 0.1 とする
- ・ 目的ダミー変数については、N=1,000、n=100、p=0.5 として分散は式(10)から算出する

結果得られた分散共分散行列は表-5 の通り。

表-5 本分析で設定された説明変数の分散共分散行列 ( $\Sigma_{ij}$ )

	Air 時間	Air 費用	Air 目的D	Air 0	Air 定数項	Air 0	Rail 時間	Rail 費用	Rail 0	Rail 目的D	Rail 0	Rail 定数項	Car 時間	Car 費用	Car 0	Car 0	Car 0	Car 0
Air 時間	0.034	0.003	0.000	0.000	0.000	0.000	0.033	0.002	0.000	0.000	0.000	0.000	0.070	0.002	0.000	0.000	0.000	0.000
Air 費用	0.003	0.001	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000
Air 目的D	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Air 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Air 定数項	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Air 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rail 時間	0.033	0.006	0.000	0.000	0.000	0.000	0.168	0.005	0.000	0.000	0.000	0.000	0.253	0.010	0.000	0.000	0.000	0.000
Rail 費用	0.002	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000
Rail 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rail 目的D	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rail 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rail 定数項	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Car 時間	0.070	0.013	0.000	0.000	0.000	0.000	0.253	0.011	0.000	0.000	0.000	0.000	0.472	0.018	0.000	0.000	0.000	0.000
Car 費用	0.002	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.018	0.001	0.000	0.000	0.000	0.000
Car 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Car 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Car 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Car 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

プログラムは Gauss を用いた。以下がそのソースコードである。

```
new , 200000;
screen on;

output file=g:\%kamiya%\agg-bias.out reset;
output on;
load beta[6,1] = g:\%kamiya%\beta.csv ;
load x[6,3] = g:\%kamiya%\x.csv ;
load sbeta[6,6] = g:\%kamiya%\sbeta.csv ;
```

```

load xs0[6*3,6*3] = g:¥kamiya¥sx.csv ;

v = x' * beta;
p = exp(v)/sumc(exp(v));

cv1 = x'*sbeta*x;

cv2=zeros(3,3);
i=1; do while i<= 3;
  j=1; do while j<=3;
    xs=xs0[ ((i-1)*6+1):i*6 , ((j-1)*6+1):j*6 ];
    cv2[i,j] = beta'*xs*beta;
  j=j+1; endo;
i=i+1; endo;

print cv1;
print cv2;

cov=cv1+cv2;

varp=zeros(3,1);
i=1; do while i <= 3;
  j=1; do while j <= 3;
    if j==i;
      a1=p[i]*(1-p[i]);
    else;
      a1=-p[i]*p[j];
    endif;
    k=1; do while k <= 3;
      if k==i;
        a2=p[i]*(1-p[i]);
      else;
        a2=-p[i]*p[k];
      endif;
      varp[i] = varp[i] + a1*a2*cov[j,k];
    k=k+1; endo;
  j=j+1; endo;
i=i+1; endo;

print sqrt(varp)~p;
print (sqrt(varp)./p);

output off;

```

パラメータ誤差および説明変数に起因する、選択肢間分散共分散行列（プログラム中では各々cv1, cv2）は下記の通り。

パラメータ誤差に起因する選択肢間効用分散共分散行列：  $\Lambda_1$

0.028727705	0.030578097	0.039121480
0.030578097	0.034132291	0.043081527
0.039121480	0.043081527	0.057936622

説明変数誤差に起因する選択肢間効用分散共分散行列：  $\Lambda_2$

0.058270207	0.067987120	0.11797078
0.067987120	0.18432177	0.29395447
0.11797078	0.29395447	0.53640028

選択肢間効用の分散共分散行列： $\text{cov}[V_i, V_j]$

0.080652356	0.63639185
0.066632662	0.32782884
0.017765488	0.035779311

これを用いた各選択肢誤差分散と推計選択シェア：

選択肢誤差分散	選択肢シェア
0.080652356	0.63639185
0.066632662	0.32782884
0.017765488	0.035779311

結果として得られる各選択肢の変動係数は以下の通り。

0.12673380  
0.20325443  
0.49652963

## 5. おわりに

今回、たまたま RM 分析に関連して、再び非集計モデルの集計誤差を整理することになった。最初に書いたとおり、身近に関連する分かりやすい資料を見いだせなかったことから、本稿執筆の運びとなった。もし本稿に類した資料をご存じであれば、是非ともご一報頂きたい。

さて、実は 1984 年 2 月提出の兵藤の卒業論文も、扱うべき題材は非集計モデルの誤差解析に関わるテーマであった。しかし前橋・高崎地区の LOS 作成プログラムに時間をとられ、実際にパラメータ誤差などを計算したのは当時修士 1 年の竹内氏であった(兵藤卒論のその章は竹内氏が計算・執筆していた)。今回、22 年ぶりに卒論時の借金をわずかなりとも返却できたような気がする。

なお、式展開など誤差の基本的考え方についてコメント頂いた福田大輔氏(東京工業大学)に謝意を表す。

以上

## <参考文献>

- 1) Ben-Akiva, M., and Lerman, S. (1985): Discrete Choice Analysis: Theory and Application to predict travel demand, MIT press
- 2) Koppelman, F.S. (1976): Methodology for Analysis Errors in Prediction With Disaggregate Choice Models, TRR?
- 3) 桐越他(1983): シミュレーションデータによる非集計ロジットモデルの誤差解析, 第 5 回土木計画学研究発表会講演集
- 4) 桐越他(1983): 非集計ロジットモデルのアグリゲーションバイアスに関する研究, 第 38 回年次学術講演会
- 5) 森地・屋井・竹内(1985): 非集計行動モデル構築に要する交通サービス特性データの精度に関する検討, 第 40 回年次学術講演会