

LEM での Log-Multiplicative Model のプログラムの書き方*

太郎丸 博†

1 はじめに

この小論では、LEM での Log-Multiplicative Model のプログラムの書き方について、簡単な例を書いてみます。LEM はカテゴリカル・データ分析のための汎用性の高いフリー・ウェアです。LEM のプログラムの書き方は、マニュアルを読めば書いてありますし、拙著でも触れてあります。しかし、LEM は柔軟性が高く、機能も豊富なせいで、プログラムの書き方はやや複雑です。マニュアルも英語ですし、一般化線形モデルを理解していることを前提に書かれているので、ここではごく簡単な例をあげてみましょう。なお、LEM は、下記の URL から入手してください。

<http://www.uvt.nl/faculiteiten/fsw/organisatie/departementen/mto/software2.html>

2 対連関モデルからいちばん簡単な Log-Multiplicative Model へ

LEM のサンプルプログラム Xie92b_1 のデータを使います。LEM のメニューからは、[Examples] [General Log-Linear] [Multiplicative Scaling] [Full Interaction] で開けます。

このデータを使って、もう少し単純なプログラムを書いたのが、下のサンプルです。これは、国 (C) × 父職 (F) × 息子職 (S) のクロス表に関して、CF と CS の full-interaction を仮定し、FS については、準独立モデルをあてはめたものです。モデルを式で書けば、

$$\log n_{ijk} = \lambda + \lambda_i^C + \lambda_j^F + \lambda_k^S + \lambda_{ij}^{CF} + \lambda_{ik}^{CS} + \tau_{jk}^{FS}$$

となります。ただし、 τ_{jk}^{FS} は主対角線上以外は必ず 0 に固定します。

```
man 3
```

```
dim 3 5 5
```

```
lab C F S *C: Country, F: Father's Job, S: Son's Job
```

```
mod {CF CS spe(FS,5a)}
```

* Wiki に 2006/08/08 に掲載したものを、Wiki サイトの閉鎖に伴い、pdf 版に変更し、内容も微修正した。

† 京都大学文学研究科, tarohmaru.h@hs2.ecs.kyoto-u.ac.jp.

```

dat [
1275 364 274 272 17 * アメリカの移動表
1055 597 394 443 31
1043 587 1045 951 47
1159 791 1323 2046 52
666 496 1031 1632 646

474 129 87 124 11 * イギリスの移動表
300 218 171 220 8
438 254 669 703 16
601 388 932 1789 37
76 56 125 295 191

127 101 24 30 12 * 日本の移動表
86 207 64 61 13
43 73 122 60 13
35 51 62 66 11
109 206 184 253 325]

```

ただ、これでは、あてはまりが非常に悪いので、Log-Multiplicative Model をあてはめてみます。つまり、

$$\log n_{ijk} = \lambda + \lambda_i^C + \lambda_j^F + \lambda_k^S + \lambda_{ij}^{CF} + \lambda_{ik}^{CS} + \tau_{jk}^{FS} \phi_i^C$$

というモデルです。これは、父職 (F) と息子職 (S) の連関については、準独立モデルがどの国に關してもあてはまるが、その連関の強さだけが、国によって異なるということを仮定します。この国による FS の連関の強さを示すパラメータが ϕ_i^C です。LEM では、最初のカテゴリのパラメータを 1 に固定し、その他のカテゴリが最初のカテゴリに比してどの程度連関が強いかを推定します。例えば、この例では、3 カ国のデータですが、最初が米国、次が英国、最後が日本ですから、米国に關しては、 $\phi_i^C = 1$ に固定され、英国と日本に關してはデータから ϕ_i^C を推定します。

このための LEM のプログラムは下記の通りです。最初のプログラムと違うのは、spe(FS,5a, C, b) というところだけです。spe(FS,5a というのは、FS に關して、準独立モデルを当てはめなさいという指定で、最初のプログラムと同じです。C は、準独立モデルの連関の強さは、C のカテゴリによって変化しますよ、ということ指定します。最後の b が simple heterogeneous な連関を仮定しているということです。詳しくはまた書きます。

```

man 3
dim 3 5 5
lab C F S *C: Country, F: Father's Job, S: Son's Job

```

```
mod {CF CS spe(FS,5a, C, b)}
dat [
データは同じなので省略]
```

このように指定すると、output では、パラメータの最後に下記のように出てきます。最初の C [spe(FS,5a)] が ϕ_i^C の推定値で、spe(FS,5a) [C] が τ_{jk}^{FS} の推定値です。

```
C [spe(FS,5a)]
  1          1.0000
  2          1.2797
  3          0.9656

spe(FS,5a) [C]
  1          1.2732      3.5721
  2          0.5086      1.6629
  3          0.2253      1.2527
  4          0.3595      1.4326
  5          2.5657     13.0092
```

ちなみに、この場合、C のカテゴリによって、FS の連関の強さが変わると仮定したわけですが、この場合の C を LEM では "group margin" と呼んでいます。

3 もうちょっと複雑な Log-Multiplicative model

例えば、父職と息子職の連関が、国の脱産業化得点 (P_i) に比例して変化するというモデルを作りたいとしましょう。そこで次のようなモデルを考えます。

$$\log n_{ijk} = \lambda + \lambda_i^C + \lambda_j^F + \lambda_k^S + \lambda_{ij}^{CF} + \lambda_{ik}^{CS} + \tau_{jk}^{FS}(\beta_0 + \beta_1 P_i)$$

例えば、脱産業化得点が、米国 0.5、英国 0.8、日本 0.4 だったとすると、プログラムの書き方は、下記の通り。プログラムの違いの 1 点目は、spe(FS,5a, C, b, -2) の部分。最後の "-2" の "2" は「des[] で指定するようなパラメータを 2 つ推定せよ」という意味です。"-2" の "-" は「des[] 中の数値は間隔尺度である」という意味です。des[] 中の指定の仕方については、デザイン行列の指定になれていないとわかりにくいかもしれませんが、それについては、別の文献をあたってください。

```
man 3
dim 3 5 5
lab C F S *C: Country, F: Father's Job, S: Son's Job
mod {CF CS spe(FS,5a, C, b, -2)}
des [1 1 1          * beta0=intercept
     0.5 0.8 0.4]  * beta1=slope of P
```

```
dat [
```

```
データは同じなので省略]
```

output は以下の通り。C [spe(FS,5a)] の 1 が $\beta_0 = 1.0000$ であることを示しています。ただし、LEM は自動的に最初のパラメータを 1 に固定するようになっています。そして、C [spe(FS,5a)] の 2 が $\beta_1 = 1.5256$ であることを示しています。つまり、脱産業化得点 (P) が大きい国ほど、父職と息子職の連関が強まるという推定結果です。

```
C [spe(FS,5a)]
```

```
1          1.0000
```

```
2          1.5256
```

```
spe(FS,5a) [C]
```

```
1          0.7288      2.0726
```

```
2          0.2909      1.3377
```

```
3          0.1283      1.1369
```

```
4          0.2057      1.2284
```

```
5          1.4758      4.3746
```

4 association model での log-multiplicative model

4.1 対連関型の連関

association model にもいろいろありますが、次のようなモデルを考えます。

$$\log n_{ijk} = \lambda + \lambda_i^C + \lambda_j^F + \lambda_k^S + \lambda_{ij}^{CF} + \lambda_{ik}^{CS} + \phi \mu_j^F \mu_k^S$$

パラメータ同士の積がモデルに組み込まれているので、これも log-multiplicative model の一種です。このためのプログラムは、下記の通り。"ass2" が log-multiplicative association の指定で、つぎに、association 型の連関を仮定する変数を指定し (この場合は F と S)、最後の "5a" の "5" は Row and Column Association の指定、"a" は変数 F と S の連関が、第 3 変数 (C) によって変化しないこと指定しています。

```
man 3
```

```
dim 3 5 5
```

```
lab C F S *C: Country, F: Father's Job, S: Son's Job
```

```
mod {CF CS ass2(F,S, 5a)}
```

```
des [1 1 1          * beta0=intercept
```

```
      0.5 0.8 0.4] * beta1=slope of P
```

```
dat [
```

```
データは同じなので省略]
```

すると、output には、

```
type 2 association (row=F column=S)
association      4.2466
row              -0.6611 -0.2902  0.0769  0.2245  0.6499
adj row         -1.3623 -0.5980  0.1584  0.4626  1.3394
column          -0.4687 -0.3046 -0.0777  0.0258  0.8252
adj column      -0.9658 -0.6277 -0.1602  0.0532  1.7004
```

と出てきます。association という項が、F と S の連関の強さを示します。

4.2 連関の強さの変化を仮定する

上の連関 (association) の強さが国によって違うと仮定する場合、例えば、次のようなモデルが考えられます。

$$\log n_{ijk} = \lambda + \lambda_i^C + \lambda_j^F + \lambda_k^S + \lambda_{ij}^{CF} + \lambda_{ik}^{CS} + \phi_i^C \mu_j^F \mu_k^S$$

このモデルでは、連関の強さ ϕ_i^C が国によって変化することを仮定します。この場合のプログラムは、下記の通り。log-multiplicative association を仮定する 2 変数 (F, S) のあとに、 ϕ_i^C を変化させる第 3 変数 (C) を指定します。そして、最後に、“5b”の”b”は単純に ϕ_i^C だけが、C にしたがって変化することを示します。

```
man 3
dim 3 5 5
lab C F S *C: Country, F: Father's Job, S: Son's Job
mod {CF CS ass2(F,S, C, 5b)}
dat [データは同じなので省略]
nco * 条件付確率の出力を省略
```

このプログラムの出力の一部が下記の通り。association のパラメータが 3 つになっていますが、これは、3 カ国それぞれについて、連関の強さを推定しているからです。

```
type 2 association (group=C row=F column=S)
association      4.1115  5.6207  2.9946
row              -0.6499 -0.2964  0.0692  0.2146  0.6625
adj row         -1.3178 -0.6009  0.1403  0.4351  1.3434
column          -0.4616 -0.3063 -0.0797  0.0191  0.8285
adj column      -0.9360 -0.6210 -0.1616  0.0387  1.6799
```