下三角行列の対数線形モデル
—支持政党と株取引の再分析—*

太郎丸 博†

Bruce G. Carruthers, 1996, *City of Capital: Politics and Markets in the English Financial Revolution*, Princeton U. P. のデータの再分析。この本では、対数線形モデルを使った分析がしてあるのだが、どのようなモデルなのか明示されていない。再分析してみると、著者はけっこうイレギュラーな分析をしていることに気がついた。分析の詳細は下記の通り。

1　問題とデータ

1712 年のロンドン証券取引所では、支持政党の同じ人と株の取引をしたがる傾向があったかどうかが、この著書では問題になっている。データは表 1 の通り。Active Traders とは、取引が相対的に多い人のこと、Inactive Traders とはその逆。Active Traders どうしの取引に限れば、Tories 党の支持者どうしの取引が 26 件、Whigs 党の支持者と Tories 党の支持者の取引が 178 件という具合に見ていく。

表 1　1712 年のロンドン証券取引所の支持政党別の取引数

Active Traders			
	Tories	Whigs	Unknown
Tories	26		
Whigs	178	496	
Unknown	77	445	93
Inactive Traders			
	Tories	Whigs	Unknown
Tories	13		
Whigs	43	69	
Unknown	57	165	135

* Wiki に 2006/07/13 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。

表 2　Inactive Traders のランダムマッチングの組み合わせ数

	Tories	Whigs	Unknown	計
Tories	$126(126-1)/2 = 7875$			
Whigs	$126 \times 346 = 43596$	$346(346-1)/2 = 59685$		
Unknown	$126 \times 492 = 61992$	$346 \times 492 = 170232$	$492(492-1)/2 = 120786$	
計				464166

1.1　独立モデルの組み方

同じ政党どうしの取引が多ければ、主対角線上のセル度数が、独立モデル（著者は準独立モデルと呼んでいる）よりも多くなるはずである。ただし、普通のクロス表と違うので、通常の独立モデルとは少し違った工夫が必要である。基本的には周辺度数から、セル度数を予測していくが、考え方がやや異なる。Inactive Traders のデータで考えてみよう。取引は合計 482 回成立しているので、取引をした延べ人数は、

$$482 \times 2 = 964$$

である。独立モデルとは、この 964 人がまったくランダムに組み合わされた場合を予測するものである。そこで、Tories と Whigs と Unknown の人数を計算してみよう。Tories の人数は、

$$13 + 13 + 43 + 57 = 126$$

である。13 を 2 回足すのは、Tories と Tories の取引には、2 人の Tories がかかわっているからである。同様にして、Whigs と Unknown の人数は

$$\text{Whigs :} \quad 43 + 69 + 69 + 165 = 346$$

$$\text{Unknown :} \quad 57 + 165 + 135 + 135 = 492$$

である。すべての人の組み合わせが同じ確率で生じると考えるのが独立モデルの自然な考え方であろう。そこで組み合わせを計算すると、表 2 のようになる。これから、各セルの確率を推定すると、下記のようになる。

	Tories	Whigs	Unknown	計
Tories	0.017			
Whigs	0.094	0.129		
Unknown	0.134	0.367	0.260	
計				1

上の確率に、実際に成立した総組み合わせ数 482 をかけると期待度数が推定できる。下記がその期待度数である。

	Tories	Whigs	Unknown	計
Tories	8.19			
Whigs	45.31	62.18		
Unknown	64.59	176.89	125.32	
計				482

この期待度数と実際の度数の標準残差を計算すると、表 3 のようになる。

確かに、対角セルの残差が大きく出ている。これから、尤度比等計量を計算すると、

$$G^2 = 5.68$$

で、著者の分析結果とほぼ一致する。自由度も 3 で一致する。

ちなみにこの独立モデルを式で表すと、Tories, Whigs, Unkonwn ののべ人数を n_T, n_W, n_U とし、Tories と Tories の取引の期待度数を E_{TT} と表記すると、

$$E_{TT} = \frac{n_T(n_T-1)}{2} \times \frac{482}{464166}$$

$$\log E_{TT} = \log[n_T(n_T-1)/2 \times \frac{482}{464166}] = \log n_T + \log(n_T-1) - \log 2 + \log 482 - \log 464166$$

となる。$\log n_T \simeq \log(n_T-1)$ とし、$\log 482 - \log 464166 = C$ とすると、

$$\log E_{TT} = 2\log n_T - \log 2 + C$$

となる。同様にして、Tories と Whigs の取引数の期待度数は、以下のように推定できる。

$$E_{TW} = n_T n_W \times \frac{482}{464166}$$

$$\log E_{TW} = \log[n_T n_W \times \frac{482}{464166}] = \log n_T + \log n_W + C$$

基本的には、のべ人数の対数と定数 C の和として、モデルを表現できるので、通常の対数線形モデルのデザイン行列で表現可能である。すなわち、

$$\log E_{ij} = \lambda + \lambda_i + \lambda_j + \lambda_{ij}$$

ただし、

$$i = j \implies \lambda_i = \lambda_j$$

表 3　独立モデルの標準残差

	Tories	Whigs	Unknown
Tories	1.68		
Whigs	−0.34	0.87	
Unknown	−0.94	−0.89	0.86

表 4　R の glm で計算するためのデータとデザイン行列

Count	Tories	Whigs	Unknown	Active	diagonal
26	2	0	0	1	1
178	1	1	0	1	0
77	1	0	1	1	0
496	0	2	0	1	1
445	0	1	1	1	0
93	0	0	2	1	1
13	2	0	0	0	1
43	1	1	0	0	0
57	1	0	1	0	0
69	0	2	0	0	1
165	0	1	1	0	0
135	0	0	2	0	1

$$\sum \lambda_i = 0$$

$$\lambda_{ij} = \begin{cases} -\log 2 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

という制約をつける。主効果と周辺度数の効果を 2 つ推定するので推定するパラメータ数は 3 つ、セルが 6 つなので、このモデルの検定の自由度は $6 - 3 = 3$ となる。ただ、対角線上だけは、$-\log 2$ というのがあって、パラメータの推定のときに邪魔になる。そこで、R の offset 関数を使ってこの項を処理することにする。

1.2　デザイン行列

これ以上は、手計算では難しいので、R2.3.0 の glm で計算する。データとデザイン行列は表 4 の通り。おおむね著者と同じものと思われる。

2　プログラムと分析結果

R 2.3.0 でのプログラムは下記の通り。

```
SE<-read.table("StockExchange.txt",header=T) #上記のデータファイルの読み込み
SE$log2<-SE$diagonal*(-log(2)) #対角セルの -log2 の項をデータに追加
attach(SE)
```

se0<-glm(Count~Tories+Whigs+Active+offset(log2), family=poisson) #offset **は傾きを 1 に固定**

se1<-glm(Count~(Tories+Whigs)*Active+offset(log2), family=poisson)

se2<-glm(Count~(Tories+Whigs)*Active+diagonal+offset(log2), family=poisson)

se3<-glm(Count~(Tories+Whigs+diagonal)*Active+offset(log2), family=poisson)

summary(se3)

anova(se0,se1,se2,se3)

detach(SE)

分析結果は下記の通り。

Model 1: Count ~ Tories + Whigs + Active + offset(log2) **主効果のみ**

Model 2: Count ~ (Tories + Whigs) * Active + offset(log2) **支持政党とトレーダーのアクティブさの交互作用効果**

Model 3: Count ~ (Tories + Whigs) * Active + diagonal + offset(log2) **対各セルの効果を追加**

Model 4: Count ~ (Tories + Whigs + diagonal) * Active + offset(log2) **対角セルとアクティブさの交互作用も追加**

```
  Resid. Df Resid. Dev Df Deviance
1         8    215.568
2         6     10.012  2  205.556
3         5      7.795  1    2.216
4         4      5.483  1    2.313
```

仮に、同じ政党どうしのほうが株取引が成立しやすいならば、対角セルに有意な効果があるはずである。つまり、上記の Model 2 と Model 3 を比べると、Model 3 のほうが、適合度がよいはずである。しかし、そうはなっていない。原著では、active traders と inactive traders を別々に分析しているということが影響しているのであろう。原著では、active traders は、同じ政党支持者どうしで取引しあう傾向は見られないが、inactive traders はその傾向があるとしている。

そこで、Model 4 の結果をしめすと、次のようになる。

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      5.39692    0.09828  54.913  < 2e-16 ***
Tories          -1.28510    0.10409 -12.346  < 2e-16 ***
Whigs           -0.32330    0.06845  -4.723 2.32e-06 ***
diagonal         0.20591    0.09792   2.103 0.035477 *
Active          -0.14889    0.12061  -1.235 0.217000
Tories:Active    0.45351    0.12486   3.632 0.000281 ***
```

```
Whigs:Active      1.13905     0.08484   13.426   < 2e-16 ***
diagonal:Active -0.17989     0.11815   -1.523 0.127865
```

この結果を見ると、確かに、主対角線の効果 diagonal は 5% 水準で有意な効果を持っている。しかし diagonal:Active（両者の交互作用効果）はマイナスになっており、著者の主張にそう結果である。ただし、この交互作用効果は有意ではないし、AIC でみても model 2, 3, 4 は適合度に大差はない。つまり、玉虫色のはっきりしない結果になっている。BIC でみれば、確実に model 2 があてはまりがよい。というわけで、著者の主張はかなり危ういことがわかる。