層別化された多群比率モデルの統計解析法

2014SS038 小森有真　2014SS058 中塩彩那

指導教員：白石高章

1　はじめに

統計学の基礎となる確率や事象を学び, それを基に統計学の分布論や二項分布に関係した 1,2 標本モデルを統計的解析法, 分散安定化変換等を学んだ. 統計学が実際にどのように用いられているかが明確になるにつれて, カイ自乗型検定に興味を持ち, この検定法について研究することにした. そこで本論では, 層別化された多群比率モデルの統計解析法について考察する.

2　カイ自乗型検定

標本の観測値を X_{ijk} $(k=1,\cdots,\ n_j\ ;\ j=1,\cdots,\ J\ ;\ i=1,\cdots,\ I)$ とし X_{ijk} は互いに独立と仮定する. また, X_{ijk} は成功の確率が p_{ij} のベルヌーイ試行とする. すなわち, $X_{ijk}\sim B(1,p_{ij})$ である. 帰無仮説 $H_0: p_{i1}=\cdots=p_{iJ}\ (i=1,\cdots,I)$ vs. 対立仮説 H_1 : ある j と j' に対して $p_{ij}\neq p_{ij'}\ (i=1,\cdots,I)$ を考える. 以下は i 層目のモデルとする $(i=1,\cdots,\ I)$. (表 1 参照)

表 1　モデル

標本	サイズ	データ	$X_{ij\cdot}$ の分布
第 1 標本	n_1	$X_{i11},\cdots,\ X_{i1n_1}$	$B(n_1,\ p_{i1})$
第 2 標本	n_2	$X_{i21},\cdots,\ X_{i2n_2}$	$B(n_2,\ p_{i2})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
第 j 標本	n_j	$X_{ij1},\cdots,\ X_{ijn_j}$	$B(n_j,\ p_{ij})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
第 J 標本	n_J	$X_{iJ1},\cdots,\ X_{iJn_J}$	$B(n_J,\ p_{iJ})$

i 層目の標本サイズ： $n\equiv n_1+\cdots+n_J$
以後, 次の (条件 1) を仮定する.

(条件 1)
$$\lim_{n\to\infty}\frac{n_j}{n}=\lambda_j>0\ \ (j=1,\cdots,\ J)$$

このとき,

$$\hat{p}_{ij}(1)\equiv\frac{1}{n_j}\sum_{k=1}^{n_j}X_{ijk}$$

$$\hat{p}_{ij}(2)\equiv\frac{1}{n_j+1}\left(\sum_{k=1}^{n_j}X_{ijk}+0.5\right)$$

とおくと, p_{ij} の点推定量は $\hat{p}_{ij}\equiv\hat{p}_{ij}(1)$, $\hat{p}_{ij}(2)$ で与えられる.
帰無仮説 H_0 の下で p_{ij} の点推定量 $\hat{p}_{ij}(1)$ の平均と分散を求める.
文献 [1] の定理 7.3 より

$$E(\hat{p}_{ij}(1))=p_{ij}\qquad(j=1,\cdots,\ J)\tag{1}$$

$$V(\hat{p}_{ij}(1))=\frac{1}{n_j}p_{ij}(1-p_{ij})\qquad(j=1,\cdots,\ J)\tag{2}$$

を得る.

【命題 2.1】
(条件 1) のもとで, $n\to\infty$ として,

$$\sqrt{n}(\hat{p}_{ij}(m)-p_{ij})\xrightarrow{\mathcal{L}}N\left(0,\frac{1}{\lambda_j}p_{ij}(1-p_{ij})\right)\qquad(m=1,2)$$

が成り立つ.

【証明】次のように場合分けをして考える.
(i)　$m=1$ のとき
(1) , (2) , 中心極限定理より,

$$\frac{\sqrt{n_j}(\hat{p}_{ij}(1)-p_{ij})}{\sqrt{p_{ij}(1-p_{ij})}}\xrightarrow{\mathcal{L}}W_{ij}\sim N(0,1)\tag{3}$$
$$(i=1,\cdots,\ I\ ;\ j=1,\cdots,\ J)$$

が成り立つ.
(3) より, $\sqrt{n}(\hat{p}_{ij}-p_{ij})$ を式変形させると以下のようになり,

$$\sqrt{n}(\hat{p}_{ij}(1)-p_{ij})=\frac{\sqrt{n}}{\sqrt{n_j}}\cdot\sqrt{p_{ij}(1-p_{ij})}\cdot\frac{\sqrt{n_j}(\hat{p}_{ij}(1)-p_{ij})}{\sqrt{p_{ij}(1-p_{ij})}}\tag{4}$$

文献 [1] のスラツキーの定理より,

$$((4)\text{ 式の右辺})\xrightarrow{\mathcal{L}}\frac{1}{\sqrt{\lambda_j}}\sqrt{p_{ij}(1-p_{ij})}W_{ij}\sim N\left(0,\frac{1}{\lambda_j}p_{ij}(1-p_{ij})\right)$$
$$(i=1,\cdots,\ I\ ;\ j=1,\cdots,\ J)$$

となる.

(ii)　$m=2$ のとき
$\sqrt{n}(\hat{p}_{ij}-p_{ij})$ を式変形させると以下のようになり,

$$\sqrt{n}(\hat{p}_{ij}(2)-p_{ij})$$
$$=\frac{\sqrt{n}}{\sqrt{n_j}}\cdot\sqrt{p_{ij}(1-p_{ij})}\cdot\frac{n_j}{n_j+1}\cdot\left\{\frac{\sqrt{n_j}\left(\hat{p}_{ij}(1)-p_{ij}\right)}{\sqrt{p_{ij}(1-p_{ij})}}\right\}+\frac{\sqrt{n}(0.5-p_{ij})}{n_j+1}\tag{5}$$

文献 [1] のスラツキーの定理より,

$$((5)\text{ 式の右辺})\xrightarrow{\mathcal{L}}\frac{1}{\sqrt{\lambda_j}}\cdot\sqrt{p_{ij}(1-p_{ij})}\cdot W_{ij}\sim N\left(0,\frac{1}{\lambda_j}p_{ij}(1-p_{ij})\right)$$

$$(i=1,\cdots,\ I\ ;\ j=1,\cdots,\ J)$$

となる．

$\square$

ここで, 文献 [1] の定理 3.35 のデルタ法を用いると,

$$2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\arcsin\left(\sqrt{p_{ij}}\right)\right\}$$
$$\xrightarrow{\mathcal{L}}\frac{1}{\sqrt{\lambda_j}}\frac{1}{\sqrt{p_{ij}(1-p_{ij})}}\sqrt{p_{ij}(1-p_{ij})}W_{ij}$$
$$=\frac{1}{\sqrt{\lambda_j}}W_{ij}$$

$$(i=1,\cdots,\ I\ ;\ j=1,\cdots,\ J)$$

が成り立ち，$Y_{ij}\equiv\dfrac{1}{\sqrt{\lambda_i}}W_{ij}$ とすると,

$$Y_{ij}\sim N\left(0,\frac{1}{\lambda_j}\right)$$

となる．

3　提案する検定法

$S_{ij}\equiv 2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\arcsin(\sqrt{p_{ij}})\right\}$ とすると,

$$S_{ij}\xrightarrow{\mathcal{L}}Y_{ij}\sim N\left(0,\frac{1}{\lambda_j}\right)\tag{6}$$

となり, 文献 [1] の定理 3.39 を用いると,

$$\sum_{j=1}^{J}c_jS_{ij}\xrightarrow{\mathcal{L}}\sum_{j=1}^{J}c_jY_{ij}\tag{7}$$

となる．

【補題 3.1】　$\hat{Z}_{ij}$ を

$$\hat{Z}_{ij}\equiv 2\sqrt{n_j}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\sum_{k=1}^{J}\frac{n_k}{n}\arcsin\left(\sqrt{\hat{p}_{ik}}\right)\right\}$$

とする. このとき，H_0 の下で $n\to\infty$ として,

$$\hat{Z}_{ij}\xrightarrow{\mathcal{L}}\sqrt{\lambda_j}\left(Y_{ij}-\sum_{k=1}^{J}\lambda_kY_{ik}\right)$$

$$(i=1,\cdots,\ I\ ;\ j=1,\cdots,\ J)$$

が成り立つ．

【証明】H_0 の下より,

$$\hat{Z}_{ij}=\frac{\sqrt{n_j}}{\sqrt{n}}\Bigg[2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\arcsin(\sqrt{p_{ij}})\right\}$$
$$-\sum_{k=1}^{J}\frac{n_k}{n}2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ik}}\right)-\arcsin(\sqrt{p_{ik}})\right\}\Bigg]$$

と式変形できる．

$S_{ij}=2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\arcsin(\sqrt{p_{ij}})\right\}$ より,

$$\hat{Z}_{ij}=\frac{\sqrt{n_j}}{\sqrt{n}}\left(S_{ij}-\sum_{k=1}^{J}\frac{n_k}{n}S_{ik}\right)$$

が成り立ち，(6)，(7) より,

$$\hat{Z}_{ij}\xrightarrow{\mathcal{L}}\sqrt{\lambda_j}\left(Y_{ij}-\sum_{k=1}^{J}\lambda_kY_{ik}\right)$$

を得る．

$\square$

検定統計量 T_B を求める前に i に対しての統計量 T_{iB} を求める．

【定理 3.2】　統計量 T_{iB} を $T_{iB}\equiv\sum_{j=1}^{J}(\hat{Z}_{ij})^2$ とすると，H_0 の下で $n\to\infty$ として,

$$T_{iB}\xrightarrow{\mathcal{L}}T\sim\chi^2_{J-1}\quad(i=1,\cdots,\ I)$$

が成り立つ．

このとき，T は文献 [1] の定理 3.23 より,

$$T\equiv\sum_{i=1}^{k}\lambda_i\left(Y_i-\sum_{j=1}^{k}\lambda_jY_j\right)^2\sim\chi^2_{k-1}\tag{8}$$

である．

【証明】

$$\hat{Z}_{ij}\ =\ 2\sqrt{n_j}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\sum_{k=1}^{J}\frac{n_k}{n}\arcsin\left(\sqrt{\hat{p}_{ik}}\right)\right\}$$

より,

$$T_{iB}=\sum_{j=1}^{J}\frac{n_j}{n}\left(S_{ij}-\sum_{k=1}^{J}\frac{n_k}{n}S_{ik}\right)^2$$

が成り立つ. 補題 3.1 より,

$$T_{iB}\xrightarrow{\mathcal{L}}\sum_{j=1}^{J}\lambda_j\left(Y_{ij}-\sum_{k=1}^{J}\lambda_kY_{ik}\right)^2$$

が成り立ち，(8) より,

$$T_{iB}\xrightarrow{\mathcal{L}}T\sim\chi^2_{J-1}$$

を得る．

$\square$

ここで,

$$T=\sum_{j=1}^{J}\lambda_j\left(W_j-\sum_{k=1}^{J}\lambda_kW_k\right)^2\sim\chi^2_{J-1}$$

$$W_j\sim N\left(\mu,\frac{1}{\lambda_j}\right)\quad(j=1,\cdots,\ J)\ \text{互いに独立}$$

となる W_j を求める.

$\hat{Z}_{\cdot j} \equiv \dfrac{1}{\sqrt{I}}\sum_{i=1}^{I}\hat{Z}_{ij}$ とすると,

$$\hat{Z}_{ij} \;=\; 2\sqrt{n_j}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\sum_{k=1}^{J}\frac{n_k}{n}\arcsin\left(\sqrt{\hat{p}_{ik}}\right)\right\}$$

より,

$$\hat{Z}_{\cdot j} = \frac{1}{\sqrt{I}}\sum_{i=1}^{I}2\sqrt{n_j}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right) -\sum_{k=1}^{J}\frac{n_k}{n}\arcsin\left(\sqrt{\hat{p}_{ik}}\right)\right\}$$

となり, H_0 の下で,

$$\hat{Z}_{\cdot j} = \frac{1}{\sqrt{I}}\sum_{i=1}^{I}\frac{\sqrt{n_j}}{\sqrt{n}}\left[2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\arcsin(\sqrt{p_{ij}})\right\} -\sum_{k=1}^{J}\frac{n_k}{n}2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ik}}\right)-\arcsin(\sqrt{p_{ik}})\right\}\right]$$

と式変形できる.
$S_{ij} = 2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\arcsin(\sqrt{p_{ij}})\right\}$ より,

$$\hat{Z}_{\cdot j} = \frac{1}{\sqrt{I}}\sum_{i=1}^{I}\frac{\sqrt{n_j}}{\sqrt{n}}\left(S_{ij}-\sum_{k=1}^{J}\frac{n_k}{n}S_{ik}\right)$$

となり, 補題 3.1 より,

$$\begin{aligned}\hat{Z}_{\cdot j} &\xrightarrow{\mathcal{L}} \frac{1}{\sqrt{I}}\sum_{i=1}^{I}\sqrt{\lambda_j}\left(Y_{ij}-\sum_{k=1}^{J}\lambda_k Y_{ik}\right)\\ &= \sqrt{\lambda_j}\left(\frac{1}{\sqrt{I}}\sum_{i=1}^{I}Y_{ij}-\sum_{k=1}^{J}\lambda_k\frac{1}{\sqrt{I}}\sum_{i=1}^{I}Y_{ik}\right) \qquad (9)\end{aligned}$$

が成り立つ. また $W_j \equiv \dfrac{1}{\sqrt{I}}\sum_{i=1}^{I}Y_{ij}$ とすると,

$$((9)\text{ 式の右辺}) = \sqrt{\lambda_j}\left(W_j-\sum_{k=1}^{J}\lambda_k W_k\right)$$

と表され, W_j は $N(0, 1/\lambda_j)$ に従う.

【定理 3.3】 統計量 T_B を $T_B \equiv \sum_{j=1}^{J}(\hat{Z}_{\cdot j})^2$ とおくと, H_0 の下で $n\to\infty$ として,

$$T_B \xrightarrow{\mathcal{L}} T \sim \chi^2_{J-1}$$

が成り立つ.

【証明】

$$\hat{Z}_{\cdot j} = \frac{1}{\sqrt{I}}\sum_{i=1}^{I}2\sqrt{n_j}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right) -\sum_{k=1}^{J}\frac{n_k}{n}\arcsin\left(\sqrt{\hat{p}_{ik}}\right)\right\}$$ より,

$$T_B = \sum_{j=1}^{J}\left[\frac{1}{\sqrt{I}}\sum_{i=1}^{I}2\sqrt{n_j}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right) -\sum_{k=1}^{J}\frac{n_k}{n}\arcsin\left(\sqrt{\hat{p}_{ik}}\right)\right\}\right]^2$$

となり, H_0 の下で,

$$T_B = \sum_{j=1}^{J}\left[\frac{1}{\sqrt{I}}\sum_{i=1}^{I}\frac{\sqrt{n_j}}{\sqrt{n}}\left[2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\arcsin(\sqrt{p_{ij}})\right\} -\sum_{k=1}^{J}\frac{n_k}{n}2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ik}}\right)-\arcsin(\sqrt{p_{ik}})\right\}\right]\right]^2$$

と式変形できる.
$S_{ij} = 2\sqrt{n}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\arcsin(\sqrt{p_{ij}})\right\}$ より,

$$T_B = \sum_{j=1}^{J}\left[\frac{1}{\sqrt{I}}\sum_{i=1}^{I}\frac{\sqrt{n_j}}{\sqrt{n}}\left(S_{ij}-\sum_{k=1}^{J}\frac{n_k}{n}S_{ik}\right)\right]^2$$

が与えられる. 補題 3.1 より,

$$\begin{aligned}T_B &\xrightarrow{\mathcal{L}} \sum_{j=1}^{J}\left[\frac{1}{\sqrt{I}}\sum_{i=1}^{I}\sqrt{\lambda_j}\left(Y_{ij}-\sum_{k=1}^{J}\lambda_k Y_{ik}\right)\right]^2\\ &= \sum_{j=1}^{J}\lambda_j\left(\frac{1}{\sqrt{I}}\sum_{i=1}^{I}Y_{ij}-\sum_{k=1}^{J}\lambda_k\frac{1}{\sqrt{I}}\sum_{i=1}^{I}Y_{ik}\right)^2 \qquad (10)\end{aligned}$$

を導くことができる. $W_j = \dfrac{1}{\sqrt{I}}\sum_{i=1}^{I}Y_{ij}$ より,

$$((10)\text{ 式の右辺}) = \sum_{j=1}^{J}\lambda_j\left(W_j-\sum_{k=1}^{J}\lambda_k W_k\right)^2 \qquad (11)$$

となる. ここで (11) 式の右辺を T とおくと, 文献 [1] の定理 3.23 より,

$$T \sim \chi^2_{J-1}$$

である.

□

よって,

$$\hat{Z}_{\cdot j} = \frac{1}{\sqrt{I}}\sum_{i=1}^{I}2\sqrt{n_j}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\sum_{k=1}^{J}\frac{n_k}{n}\arcsin\left(\sqrt{\hat{p}_{ik}}\right)\right\}$$

より,

$$T_B = \frac{4}{I}\sum_{j=1}^{J}n_j\left[\sum_{i=1}^{I}\left\{\arcsin\left(\sqrt{\hat{p}_{ij}}\right)-\sum_{k=1}^{J}\frac{n_k}{n}\arcsin\left(\sqrt{\hat{p}_{ik}}\right)\right\}\right]^2$$

と表現できる.

検定方式

ここで検定統計量 T_B に基づく漸近的な検定方式を考える. 帰無仮説 H_0 vs. 対立仮説 H_1 に対する水準 α の検定は,

$T_B \geqq \chi^2_{J-1}(\alpha)$ のとき H_0を棄却する
$T_B < \chi^2_{J-1}(\alpha)$ のとき H_0を棄却しない

で与えられる．ただし，自由度 $J-1$ のカイ二乗分布の上側 $100\alpha\%$ 点 を $\chi^2_{J-1}(\alpha)$ とする．
検定関数 ϕ(x) を使って表すと，

$$\phi(x) = \begin{cases} 1 & (T_B \geqq \chi^2_{J-1}(\alpha)) \\ 0 & (T_B < \chi^2_{J-1}(\alpha)) \end{cases}$$

である．

4 C 言語によるプログラム解説

4.1 プログラム内容

C 言語により, カイ自乗型検定による検定結果及び, カイ二乗分布に従う変数を用いた検定結果を作成した. ただし, 上側 $100\alpha\%$ 点を求めるために文献 [2] を引用した.

4.2 カイ自乗型検定の main プログラム

```
int main(void)
{
    input();
    alpha();
    Xijk_pij();
    as();
    TB1();
    TB2();
    prokai1();
    prokai2();
    fai1();
    fai2();
}
```

4.3 プログラムの流れ

1. input 関数の中で, ファイル名を入力し, 層, 標本, サイズの値を入力してデータを読み込み, 整理する．
2. alpha 関数の中で，自由度と α の値を入力する．
3. Xijk_pij 関数の中で, データを 0,1 の値で表し，$\hat{p}_{ij}(1)$, $\hat{p}_{ij}(2)$ を計算.
4. as 関数の中で，arcsin$(\sqrt{\hat{p}_{ij}})$ を計算.
5. TB1 関数の中で，$\hat{p}_{ij}(1)$ のときの統計量 T_B を計算.
6. TB2 関数の中で，$\hat{p}_{ij}(2)$ のときの統計量 T_B を計算.
7. prokai1 関数の中で，$\hat{p}_{ij}(1)$ のときのカイ自乗分布の上側確率の値を求め, 棄却するか判定して出力.
8. prokai2 関数の中で，$\hat{p}_{ij}(2)$ のときのカイ自乗分布の上側確率の値を求め, 棄却するか判定して出力.
9. fai1 関数の中で，$\hat{p}_{ij}(1)$ のときの統計量と $100\alpha\%$ 点を比べて棄却するか判定して出力.
10. fai2 関数の中で，$\hat{p}_{ij}(2)$ のときの統計量と $100\alpha\%$ 点を比べて棄却するか判定して出力.

5 データとその結果解析

5.1 データ

平成 26 年，平成 27 年，平成 28 年において, 腸管感染症による死亡者数, 結核による死亡者数, ウイルス肝炎による死亡者数, 貧血による死亡者数, 糖尿病による死亡者数, パーキンソン病による死亡者数, アルツハイマー病による死亡者数, 高血圧性疾患による死亡者数, 胃潰瘍及び十二指腸潰瘍による死亡者数, ヘルニア及び腸閉塞による死亡者数が月別に分かれたデータ (文献 [3]) をもとに, プログラムを実行した．
このとき，$I=3$(年代)，$J=10$(各死亡要因)，$k=12$(月) とする．

5.2 実行結果

$\alpha=0.01$ としたときの, プログラムの実行結果のディスプレイ上の表示が次である．

```
誤差 0.000010 以下の自由度 9.000000 のカイ自乗分布の上側 1.000000 パーセント点は 21.666039
Tb1=249.417639
自由度 9 のカイ自乗分布の上側確率： 0.000000
統計量 1 は H0 を棄却する
Tb2=160.847243
自由度 9 のカイ自乗分布の上側確率： 0.000000
統計量 2 は H0 を棄却する
```

5.3 解析結果とその考察

検定の結果, 有意水準 $\alpha=0.01$ のとき帰無仮説 H_0 は棄却された．よって, 各要因の死亡者数と月ごとの関係があることが分かった．その理由として, 暖房や冷房をつける室内と屋外, 昼夜など気温の差が大きくなる時期には体調が崩れやすく, 気温差が小さくなる時期には体調が崩れにくいと考察できる．

6 おわりに

本論では, 多群比率モデルにおけるカイ自乗型検定を提案した. また，C 言語によってプログラムを作成し結果を得ることができた．実際にプログラムを作成し現実のデータを用いることによって, 層別化された多群比率モデルの統計解析法に対する理解をより深めることができた．

参考文献

[1] 白石高章：『統計科学の基礎-データと確率の結びつきがよくわかる数理』, 日本評論社, 2012.

[2] 早川由宏：『Fortran と C 言語による統計プログラミングの基礎-Mathematica の使い方』, 南山大学大学院理工学研究科, 2015.

[3] 名古屋市役所：名古屋市統計年鑑, http://www.city.nagoya.jp/shisei/category/67-5-9-45-0-0-0-0-0-0.html
2017 年 12 月 12 日閲覧