多群指数モデルにおける一様性の仮説検定法

2015SS070 髙田秀真　2015SS077 田中友規

指導教員：白石高章

1　はじめに

統計学を学ぶ過程で, 多群指数モデルに興味を持った. そのため, 多群指数モデルを多方面に理解するために研究をすることにした. 本論では, 多群指数モデルにおける一様性の仮説検定法について考察する.

2　モデルの設定と統計量の基本的性質

表 1. k 群モデル

群	サイズ	データ	平均	分布関数
第 1 群	n_1	$X_{11},\ldots,X_{1n_1}$	μ_1	$1-e^{-x/\mu_1}$
第 2 群	n_2	$X_{21},\ldots,X_{2n_2}$	μ_2	$1-e^{-x/\mu_2}$
$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$	$\vdots$	$\vdots$
第 k 群	n_k	$X_{k1},\ldots,X_{kn_k}$	μ_k	$1-e^{-x/\mu_k}$

ある要因 A があり, k 個の水準 $A_1,\cdots,A_k$ を考える. 水準 A_i における標本の観測値 $(X_{i1},X_{i2},\cdots,X_{in_i})$ を第 i 標本とよび, 平均が μ_i である密度関数 $(1/\mu_i)e^{-x/\mu_i}(x>0)$ をもつ指数分布とする. すなわち,

$$\begin{aligned} P(X_{ij} \leqq x) &= 1-e^{-x/\mu_i} \quad (x>0)\,, \\ E(X_{ij}) &= \mu_i\,, \\ V(X_{ij}) &= \mu_i^2 \end{aligned} \tag{1}$$

である. さらにすべての X_{ij} は互いに独立であると仮定する. 総標本サイズを $n \equiv n_1+\cdots+n_k$ とおく. 第 i 群の標本平均 $\bar{X}_{i\cdot} \equiv (1/n_i)\sum_{j=1}^{n_i} X_{ij}$ が, μ_i の一様最小分散不偏推定量である. ここで, その推定量を $\hat{\mu}_i = \bar{X}_{i\cdot}$ で表記する.

このとき, 文献 [1] より, 次の命題が成り立つ.

命題 1

$$\frac{2n_i\hat{\mu}_i}{\mu_i} \sim \chi^2_{2n_i} \quad (i=1,\cdots,k)$$

となる. ただし, $Y \sim \chi^2_m$ は Y が自由度 m のカイ自乗分布に従うことを意味している.

証明　仮定から, 文献 [2] の p.111 の表現を使って

$$X_{i1},X_{i2},\ldots,X_{in} \sim EX\left(\frac{1}{\mu_i}\right) = GA\left(1,\frac{1}{\mu_i}\right)$$

文献 [2] の系 3.16 より,

$$W \equiv \sum_{j=1}^{n} X_{ij} \sim GA\left(n_i,\frac{1}{\mu_i}\right)$$

となり, 密度関数は文献 [2] の p.111 より,

$$f_W(x) = \frac{\left(\frac{1}{\mu_i}\right)^n}{\Gamma(n_i)} x^{n-1} e^{-\frac{1}{\mu_i}x}$$

となる. このとき,

$$T \equiv \frac{2n_i\hat{\mu}_i}{\mu_i}$$

とおくと,

$$T = \frac{2}{\mu_i}W$$

であるから,

$$\begin{aligned} f_T(y) &= \frac{1}{\left|\frac{2}{\mu_i}\right|} f_W\left(\frac{y-0}{\frac{2}{\mu_i}}\right) = \frac{1}{\frac{2}{\mu_i}} \frac{\left(\frac{1}{\mu_i}\right)^n}{\Gamma(n_i)} \left(\frac{y}{\frac{2}{\mu_i}}\right)^{n-1} e^{-\frac{1}{\mu_i}\frac{y}{2/\mu_i}} \\ &= \frac{1}{2^n\Gamma(n_i)} y^{n-1} e^{-\frac{y}{2}} \\ &= \left(\chi^2_{2n_i}\text{の密度関数}\right) \end{aligned}$$

が成り立つ. □

命題 2　漸近理論を述べるために,

$$(\text{条件 } 1) \qquad \lim_{n\to\infty} \frac{n_i}{n} = \lambda_i > 0 \quad (1 \leqq i \leqq k)$$

を仮定する. このとき, $n \to \infty$ として,

$$\begin{aligned} &\sqrt{n}\{\log(\hat{\mu}_i) - \log(\mu_i)\} \\ &\qquad \xrightarrow{\mathcal{L}} Y_i \sim N\left(0,\frac{1}{\lambda_i}\right) \quad (i=1,\cdots,k) \end{aligned}$$

が成り立つ. ただし, $\xrightarrow{\mathcal{L}}$ は分布 (法則) 収束を表す.

証明　(1) より

$$\begin{aligned} V(\hat{\mu}_i) &= V(\bar{X}_{i\cdot}) \\ &= V\left(\left(\frac{1}{n_i}\right)\sum_{j=1}^{n_i} \bar{X}_{ij}\right) \\ &= \frac{1}{n_i}\sum_{j=1}^{n_i} V(\bar{X}_{ij}) \\ &= \frac{1}{n_i^2} n_i \mu_i^2 \\ &= \frac{\mu_i^2}{n_i} \end{aligned}$$

であるから中心極限定理より,

$$\frac{\hat{\mu}_i-\mu_i}{\sqrt{\mu_i^2/n_i}} = \frac{\hat{\mu}_i-\mu_i}{\sqrt{V(\hat{\mu}_i)}} \xrightarrow{\mathcal{L}} U_i \sim N(0,1)$$

となる. このとき

$$\sqrt{\frac{n}{n_i}} \xrightarrow{P} \sqrt{\frac{1}{\lambda_i}}$$

であるから, スラツキーの定理より,

$$\frac{\sqrt{n}(\hat{\mu}_i - \mu_i)}{\mu_i} \xrightarrow{\mathcal{L}} \sqrt{\frac{1}{\lambda_i}} U_i \tag{2}$$

が成り立つ. ただし, $\xrightarrow{P}$ は確率収束を表す. (2) から

$$\sqrt{n}(\hat{\mu}_i - \mu_i) \xrightarrow{\mathcal{L}} \sqrt{\frac{1}{\lambda_i}} \mu_i U_i$$

が言えるので, デルタ法より

$$\begin{aligned}\sqrt{n}\{\log(\hat{\mu}_i) - \log(\mu_i)\} &\xrightarrow{\mathcal{L}} \frac{d\log(\mu_i)}{d\mu_i} \sqrt{\frac{1}{\lambda_i}} \mu_i U_i \\ &= \sqrt{\frac{1}{\lambda_i}} U_i\end{aligned}$$

が成り立つ. この平均と分散はそれぞれ

$$E\left(\sqrt{\frac{1}{\lambda_i}} U_i\right) = 0\,, \qquad V\left(\sqrt{\frac{1}{\lambda_i}} U_i\right) = \frac{1}{\lambda_i}$$

となり, $Y_i \equiv \sqrt{\frac{1}{\lambda_i}} U_i$ とすれば命題が成り立つ. □

3 一様性の検定法

水準 α の

$$\begin{cases} \text{帰無仮説 } H_0\text{:} & \mu_1 = \mu_2 = \cdots = \mu_k \\ \text{対立仮説 } H_1\text{:} & \text{ある } i\,, i' \text{ に対して } \mu_i \neq \mu_{i'} \end{cases}$$

に対する検定を考える.

$$T_i \equiv \sqrt{n}(\log \hat{\mu}_i - \log \mu_i) \tag{3}$$

とおく. このとき, 命題 2 より

$$T_i \xrightarrow{\mathcal{L}} Y_i \sim N\left(0, \frac{1}{\lambda_i}\right) \tag{4}$$

が言える.

命題 3 (条件 1) と H_0 の下で

$$\hat{Z}_i \equiv \sqrt{n_i}\left(\log \hat{\mu}_i - \sum_{j=1}^{k} \frac{n_j}{n} \log \hat{\mu}_j\right)$$

とおく. このとき, $n \to \infty$ として,

$$\hat{Z}_i \xrightarrow{\mathcal{L}} \sqrt{\lambda_i}\left(Y_i - \sum_{j=1}^{k} \lambda_j Y_j\right)$$

が成り立つ.

証明 $\hat{Z}_i$ を変形すると,

$$\begin{aligned}\hat{Z}_i &= \sqrt{n_i}\left(\log \hat{\mu}_i - \sum_{j=1}^{k} \frac{n_j}{n} \log \hat{\mu}_j\right) \\ &= \sqrt{n_i} \log \hat{\mu}_i - \sqrt{n_i} \sum_{j=1}^{k} \frac{n_j}{n} \log \hat{\mu}_j \\ &= \frac{\sqrt{n_i}}{\sqrt{n}} \left\{ \sqrt{n}\,(\log \hat{\mu}_i - \log \mu_i) - \sqrt{n} \sum_{j=1}^{k} \frac{n_j}{n} \left(\log \hat{\mu}_j - \log \mu_j\right) \right\} \\ &= \frac{\sqrt{n_i}}{\sqrt{n}} \left\{ \sqrt{n}(\log \hat{\mu}_i - \log \mu_i) - \sum_{j=1}^{k} \frac{n_j}{n} \sqrt{n}(\log \hat{\mu}_j - \log \mu_j) \right\}\end{aligned}$$

となる. ここで (3) を使ってさらに変形すると,

$$\hat{Z}_i = \frac{\sqrt{n_i}}{\sqrt{n}} \left(T_i - \sum_{j=1}^{k} \frac{n_j}{n} T_j\right)$$

となる. このことと (4) より,

$$\hat{Z}_i \xrightarrow{\mathcal{L}} \sqrt{\lambda_i}\left(Y_i - \sum_{j=1}^{k} \lambda_j Y_j\right) \tag{5}$$

が成り立つ. □

定理 1

$$T_Z \equiv \sum_{i=1}^{k} (\hat{Z}_i)^2$$

とおく. 帰無仮説 H_0 の下で, $n \to \infty$ とすると,

$$T_Z \xrightarrow{\mathcal{L}} T \sim \chi^2_{k-1}$$

が成り立つ.

証明 T_Z は,

$$\begin{aligned}T_Z &= \sum_{i=1}^{k} \left\{ \frac{\sqrt{n_i}}{\sqrt{n}} \left(T_i - \sum_{j=1}^{k} \frac{n_j}{n} Tj\right) \right\}^2 \\ &= \sum_{i=1}^{k} \frac{n_i}{n} \left(T_i - \sum_{j=1}^{k} \frac{n_j}{n} Tj\right)^2 = \sum_{i=1}^{k} \hat{Z}_i^2\end{aligned}$$

と変形でき, (5) より $T_Z \xrightarrow{\mathcal{L}} T \equiv \sum_{i=1}^{k} \lambda_i (Y_i - \sum_{j=1}^{k} \lambda_j Y_l)^2$ となる. 文献 [3] の定理 A.1 を適用して,

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

となり, 定理が成り立つ. □

$$\begin{aligned}T_Z &= \sum_{i=1}^{k} (\hat{Z}_i)^2 \\ &= \sum_{i=1}^{k} n_i \left(\log \hat{\mu}_i - \sum_{j=1}^{k} \frac{n_j}{n} \log \hat{\mu}_j\right)^2\end{aligned}$$

と書き換えられる.

T_Z の漸近分布は上式のようになる. また統計量 T_Z に基づく漸近的な検定方式について考えると, 帰無仮説 H_0 vs. 対立仮説 H_1 に対する水準 α の検定方式は

$$\begin{cases} T_Z \geq \chi^2_{k-1}(\alpha) \text{ のとき, } H_0 \text{を棄却する} \\ T_Z < \chi^2_{k-1}(\alpha) \text{ のとき, } H_0 \text{を棄却しない} \end{cases}$$

で与えられる. ただし, 自由度 $k-1$ のカイ二乗分布の上側 $100\alpha\%$ 点を $\chi^2_{k-1}(\alpha)$ とおく. このとき, 検定関数 $\phi(\boldsymbol{x})$ を使って表すと,

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & (T_Z \geq \chi^2_{k-1}(\alpha)) \\ 0 & (T_Z < \chi^2_{k-1}(\alpha)) \end{cases}$$

と表現される.

4　平均に順序制約がある場合の検定法

水準 α の

$$\begin{cases} \text{帰無仮説 } H_0: & \mu_1 = \mu_2 = \cdots = \mu_k \\ \text{対立仮説 } H_1: & \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k \\ & (\text{少なくとも 1 つの} \leq \text{は} < \text{である}) \end{cases}$$

に対する検定を考える.

H_0 が真のとき, $\mu_i = \mu_0$ とおくと,

$$\tilde{Z}_i \equiv \sqrt{n_i}\,\{\log(\hat{\mu}_i) - \log(\mu_0)\} \xrightarrow{\mathcal{L}} Z_i = \sqrt{\lambda_i} Y_i \sim N(0,1)$$

が成り立っている. このとき, 文献 [2] の定理 2.29 より, $Z_1, \ldots, Z_n$ は互いに独立である.

$$T_{\mathbf{c}} \equiv \sum_{i=1}^{k} (c_i - \bar{c})\sqrt{n_i}\sqrt{n_i}\log(\hat{\mu}_i) = \sum_{i=1}^{k}(c_i - \bar{c}) n_i \log(\hat{\mu}_i)$$

とおく. ただし, $\bar{c} = \frac{1}{n}\sum_{i=1}^{k} n_i c_i$ である.

定理 2　統計量 $\hat{T}_c$ を

$$\hat{T}_{\mathbf{c}} \equiv \frac{T_{\mathbf{c}}}{\sqrt{\sum_{i=1}^{k} n_i (c_i - \bar{c})^2}}$$

とおくと, (条件 1) と H_0 の下で, $n \to \infty$ として,

$$\hat{T}_{\mathbf{c}} \xrightarrow{\mathcal{L}} N(0,1)$$

が成り立つ.

証明

$$\begin{aligned} \sum_{i=1}^{k} n_i (c_i - \bar{c}) &= \sum_{i=1}^{k} n_i c_i - \bar{c}\sum_{i=1}^{k} n_i \\ &= \sum_{i=1}^{k} n_i c_i - \sum_{i=1}^{k} n_i c_i \\ &= 0 \end{aligned}$$

であるので,

$$T_{\mathbf{c}} = \sum_{i=1}^{k}(c_i - \bar{c}) n_i \{\log(\hat{\mu}_i) - \log(\mu_0)\} = 0$$

となる. また,

$$\bar{c}_0 \equiv \lim_{n\to\infty} \bar{c} = \sum_{i=1}^{k} \lambda_i c_i$$

とおくと, スラツキーの定理, 文献 [2] の系 3.6 より

$$\begin{aligned} \frac{T_{\mathbf{c}}}{\sqrt{n}} &= \frac{\sum_{i=1}^{k}(c_i - \bar{c})\sqrt{n_i}\sqrt{n_i}\,\{\log(\hat{\mu}_i) - \log(\mu_0)\}}{\sqrt{n}} \\ &\xrightarrow{\mathcal{L}} \sum_{i=1}^{k}(c_i - \bar{c}_0)\sqrt{\lambda_i} Z_i \qquad (6) \\ &\sim N\left(0\,, \sum_{i=1}^{k}(c_i - \bar{c}_0)^2 \lambda_i\right) \qquad (7) \end{aligned}$$

(6) を変形すると,

$$\begin{aligned} \sum_{i=1}^{k}(c_i - \bar{c}_0)\sqrt{\lambda_i} Z_i &= \sum_{i=1}^{k}\left(c_i - \sum_{j=1}^{k}\lambda_j c_j\right)\sqrt{\lambda_i} Z_i \\ &= \sum_{i=1}^{k}\sqrt{\lambda_i} c_i Z_i - \sum_{j=1}^{k}\lambda_j c_j \sum_{i=1}^{k}\sqrt{\lambda_i} Z_i \end{aligned}$$

となり, (7) より定理が成り立つ. □

ここで, $c_i = i \;\; (i = 1, \ldots, k)$ とした $\hat{T}_{\mathbf{c}}$ を $\hat{T}_k$ とすると, $\hat{T}_k$ を帰無仮説 H_0 vs 対立仮説 H_1 に対する検定統計量とすることができる. すなわち

$$\hat{T}_k = \frac{\sum_{i=1}^{k}\left(i - \frac{1}{n}\sum_{j=1}^{k} j n_j\right) n_i \log(\hat{\mu}_i)}{\sqrt{\sum_{i=1}^{k} n_i \left(i - \frac{1}{n}\sum_{j=1}^{k} j n_j\right)^2}}$$

となる. ここで検定統計量 $\hat{T}_k$ について検定方式を考える.

帰無仮説 H_0 vs 対立仮説 H_1 に対する水準 α の漸近的な検定は,

$$\begin{cases} \hat{T}_k \geq z(\alpha) \text{ のとき}, H_0 \text{を棄却する} \\ \hat{T}_k < z(\alpha) \text{ のとき}, H_0 \text{を棄却しない} \end{cases}$$

で与えられる. ただし, 標準正規分布の上側 100α %点を $z(\alpha)$ とする. 検定関数 $\phi(\boldsymbol{x})$ を使って表すと,

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & (\hat{T}_k \geq z(\alpha)) \\ 0 & (\hat{T}_k < z(\alpha)) \end{cases}$$

と表現される.

5　C 言語におけるプログラム解説

5.1　プログラムの流れ

1. 関数 input によりデータを入力
2. 関数 keisanmu によりデータのそれぞれの群の平均 μ_i を計算
3. 関数 keisanmuhat により推定量 $\hat{\mu}_i$ を計算
4. 関数 junjyotoi により出力したい結果を選択し, それを出力する

5.2　main プログラム

```
int main(void)
{
    input();
    keisanmu();
    keisanmuhat();
    junjyotoi();
    return(0);
}
```

6 C 言語プログラムによる実際のデータを用いた実行結果と解析結果

6.1 データ

熊本地震における発生した当日から震度 3 以上の地震の 1 週間ごとのデータを 3 週間とり, 第 1 群を 1 週目, 第 2 群を 2 週目, 第 3 群を 3 週目として, 第 5 節の C 言語プログラムを実行し, 6.2 節にその実行結果を述べ, 6.3 節に解析結果を論述した.

表 2. 熊本地震

群	サイズ	平均 (時)
第 1 群	98	1.414115646
第 2 群	28	5.708928571
第 3 群	23	7.402898551

6.2 実行結果

(1) $\alpha = 0.01$ のとき

```
群の個数を入力してください:3
第１群の標本数を入力してください:98
第２群の標本数を入力してください:28
第３群の標本数を入力してください:23
データのファイル名を入力してください
eqdataku.txt
0 より大きく 0.5 より小さな有効水準αの値を入力して
ください:  0.01
順序は問いますか | yes:1/no:2 |:1
誤差 0.000010 の標準正規分布の上側 1.000000 ％点
は 3.000008
帰無仮説 H_0: μ_1 = ・・・ = μ_k
対立仮説 H_1: μ_1 <= ・・・<= μ_k(少なくとも１つ
の<=は<≠である )
多群指数モデルにおける一様性の仮説検定
H_0 を棄却する

群の個数を入力してください:3
第１群の標本数を入力してください:98
第２群の標本数を入力してください:28
第３群の標本数を入力してください:23
データのファイル名を入力してください
eqdataku.txt
0 より大きく 0.5 より小さな有効水準αの値を入力して
ください:  0.01
順序は問いますか | yes:1/no:2 |:2
誤差 0.000010 の標準正規分布の上側 1.000000 ％点
は 0.580650
帰無仮説 H_0: μ_1 = ・・・ = μ_k
対立仮説 H_1: μ_1 ≠  ・・・ ≠ μ _k
多群指数モデルにおける一様性の仮説検定
H_0 を棄却する
```

(2) $\alpha = 0.05$ のとき

```
群の個数を入力してください:3
0 より大きく 0.5 より小さな有効水準αの値を入力して
ください:  0.05
順序は問いますか | yes:1/no:2 |:1
誤差 0.000010 の標準正規分布の上側 5.000000 ％点
は 3.000008
帰無仮説 H_0: μ_1 = ・・・ = μ_k
対立仮説 H_1: μ_1 <= ・・・<= μ_k(少なくとも１つ
の<=は<≠である )
多群指数モデルにおける一様性の仮説検定
H_0 を棄却する

群の個数を入力してください:3
0 より大きく 0.5 より小さな有効水準αの値を入力して
ください:  0.05
順序は問いますか | yes:1/no:2 |:2
誤差 0.000010 の標準正規分布の上側 5.000000 ％点
は 0.556740
帰無仮説 H_0: μ_1 = ・・・ = μ_k
対立仮説 H_1: μ_1 ≠  ・・・ ≠ μ _k
多群指数モデルにおける一様性の仮説検定
H_0 を棄却する
```

6.3 解析結果

α の値が 0.01 の場合と α の値が 0.05 の場合, 順序制約がある場合と, 順序制約がない場合のすべての場合で棄却された. このことから, 地震の余震の頻度は, 地震が発生してから時間がたつほど大きく下がることが分かった.

7 おわりに

本論では, 多群指数モデルにおける一様性の仮説検定法について考察した. C 言語プログラムを作成し, 現実のデータを用いることで, 多群指数モデルにおける一様性の仮説検定法に対する理解を深めることができた.

参考文献

[1] 廣瀬由幸, 森下史章 :『多群指数モデルにおけるすべての平均相違の多重比較法』 (2012)

[2] 白石高章 :『統計科学の基礎』日本評論社，東京，(2012)

[3] 白石高章 :『多群連続モデルにおける多重比較法』共立出版, 東京, (2011)

[4] 早川由宏, 白石高章 : 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. 研究ノート.（2015 年 2 月）

[5] 白石高章：多群指数モデルにおける平均パラメータの多重比較法. 計量生物学 34 巻 1-21. (2013)

[6] 気象庁：震度データベース検索. (2019 年 1 月)
http://www.data.jma.go.jp/svd/eqdb/data/shindo/index.php