指数分布モデルにおける対数対比の多重比較法

2019SS082 都竹凜花

指導教員：白石高章

1　はじめに

本研究では，漸近理論を考え，シェフェの方法を基に指数分布モデルにおける対数対比の多重比較法について考察する．そして，考察した理論に基づいて C 言語プログラムを作成し，地震の発生間隔について解析を行う．

2　モデルの設定

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	$X_{11}, \cdots, X_{1n_1}$	μ_1	$1-e^{-x/\mu_1}$
第 2 群	n_2	$X_{21}, \cdots, X_{2n_2}$	μ_2	$1-e^{-x/\mu_2}$
$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$	$\vdots$	$\vdots$
第 k 群	n_k	$X_{k1}, \cdots, X_{kn_k}$	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 標本または第 i 群とよび，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．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．

3　対数対比の多重比較法

R^k(k 次元ユークリッド空間) の部分集合 $\mathcal{C}^k$ を

$$\mathcal{C}^k \equiv \left\{ \boldsymbol{c} \equiv (c_1, \cdots, c_k) \;\middle|\; \sum_{i=1}^{k} c_i = 0\ ,\ \boldsymbol{c} \neq \boldsymbol{0} \right\}$$

で定義する．$\boldsymbol{c} \in \mathcal{C}^k$ のとき，$\sum\limits_{i=1}^{k} c_i \log \mu_i$ を対数対比といい，定数 c_i を対数対比係数と言うことにする．1 つの対数対比の仮説は，$\boldsymbol{c} \in \mathcal{C}^k$ に対して

$$\text{帰無仮説 } H_{\boldsymbol{c}}\ :\ \sum_{i=1}^{k} c_i \log \mu_i = 0$$

$$\text{vs.}\quad \text{対立仮説 } H_{\boldsymbol{c}}^A\ :\ \sum_{i=1}^{k} c_i \log \mu_i \neq 0$$

となる．すべての帰無仮説からなるファミリーを $\mathcal{F} \equiv \left\{ H_{\boldsymbol{c}} \mid \boldsymbol{c} \in \mathcal{C}^k \right\}$ で定義する．

4　漸近的な多重比較検定法

文献 [1] の命題 2 より，

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

を仮定したとき，$n \to \infty$ として，

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

が成り立つ．ただし，$\xrightarrow{\mathcal{L}}$ は分布 (法則) 収束を表す．
このことから，スラツキーの定理より，

$$c_i\sqrt{n}(\log \hat{\mu}_i - \log \mu_i) \xrightarrow{\mathcal{L}} c_i Y_i \sim N\left(0, \frac{c_i^2}{\lambda_i}\right)$$

が成り立つ．そして，文献 [2] の定理 3.39 と系 3.6 より，

$$\sum_{i=1}^{k} c_i\sqrt{n}(\log \hat{\mu}_i - \log \mu_i) \xrightarrow{\mathcal{L}} \sum_{i=1}^{k} c_i Y_i \sim N\left(0, \sum_{i=1}^{k} \frac{c_i^2}{\lambda_i}\right) \qquad (2)$$

が成り立つ．また (条件 1) より，

$$\frac{1}{\sqrt{\sum\limits_{i=1}^{k} \frac{n}{n_i} c_i^2}} \xrightarrow{P} \frac{1}{\sqrt{\sum\limits_{i=1}^{k} \frac{c_i^2}{\lambda_i}}} \qquad (3)$$

が成り立つ．ただし，$\xrightarrow{P}$ は確率収束を表す．
(2)，(3) に対してスラツキーの定理より，

$$T_{\boldsymbol{c}}(\boldsymbol{\mu}) \equiv \frac{\sum\limits_{i=1}^{k} c_i\sqrt{n}(\log \hat{\mu}_i - \log \mu_i)}{\sqrt{\sum\limits_{i=1}^{k} \frac{n}{n_i} c_i^2}} \xrightarrow{\mathcal{L}} N(0,1)$$

が成り立つ．$T_{\boldsymbol{c}}(\boldsymbol{\mu})$ の分子を $S_{\boldsymbol{c}}(\boldsymbol{\mu})$ とし，

$$\begin{aligned} &S_{\boldsymbol{c}}(\boldsymbol{\mu}) \\ &= \sum_{i=1}^{k} c_i\sqrt{n}(\log \hat{\mu}_i - \log \mu_i) \\ &= \sum_{i=1}^{k} \sqrt{\frac{n}{n_i}} c_i \\ &\quad \times \sqrt{n_i}\left\{ (\log \hat{\mu}_i - \log \mu_i) - \sum_{j=1}^{k} \frac{n_j}{n}(\log \hat{\mu}_j - \log \mu_j) \right\} \end{aligned}$$

と表現できる．ここで，

$$T_{\boldsymbol{c}} \equiv \frac{\left(\sum\limits_{i=1}^{k} c_i\sqrt{n}\log \hat{\mu}_i\right)^2}{\sum\limits_{i=1}^{k} \frac{n}{n_i} {c_i}^2} \quad , \quad T_{\boldsymbol{c}}^*(\boldsymbol{\mu}) \equiv \frac{\{S_{\boldsymbol{c}}(\boldsymbol{\mu})\}^2}{\sum\limits_{i=1}^{k} \frac{n}{n_i} {c_i}^2}$$

とおく．$H_{\boldsymbol{c}}$ の下で，$T_{\boldsymbol{c}} = T_{\boldsymbol{c}}^*(\boldsymbol{\mu})$ である．
$\{S_{\boldsymbol{c}}(\boldsymbol{\mu})\}^2$ についてシュワルツの不等式を適用すると，

$$\begin{aligned} &\{S_{\boldsymbol{c}}(\boldsymbol{\mu})\}^2 \leqq \left(\sum_{i=1}^{k} \frac{n}{n_i} {c_i}^2\right) \\ &\times \left(\sum_{i=1}^{k} n_i \left\{ (\log \hat{\mu}_i - \log \mu_i) - \sum_{j=1}^{k} \frac{n_j}{n}(\log \hat{\mu}_j - \log \mu_j) \right\}^2 \right) \end{aligned}$$

の関係が成り立つ．ゆえに，

$$T_{\boldsymbol{c}}^*(\boldsymbol{\mu}) \leqq T_0(\boldsymbol{\mu}) \equiv \sum_{i=1}^{k} n_i \left\{ (\log \hat{\mu}_i - \log \mu_i) - \sum_{j=1}^{k} \frac{n_j}{n} (\log \hat{\mu}_j - \log \mu_j) \right\}^2$$

が成り立つ．(条件 1)，(1) より，

$$T_{\boldsymbol{c}}^*(\boldsymbol{\mu}) \leqq T_0(\boldsymbol{\mu}) \xrightarrow{\mathcal{L}} \sum_{i=1}^{k} \lambda_i \left(Y_i - \sum_{j=1}^{k} \lambda_j Y_j \right)^2$$

が成り立つ．そして $\lambda_1 + \cdots + \lambda_k = 1$ より，文献 [2] の定理 3.23 を適用すると，$H_{\boldsymbol{c}}$ の下で，

$$T_{\boldsymbol{c}} = T_{\boldsymbol{c}}^*(\boldsymbol{\mu}) \leqq T_0(\boldsymbol{\mu}) \xrightarrow{\mathcal{L}} \chi_{k-1}^2 \tag{4}$$

の関係が成り立つ．ただし，χ_{k-1}^2 は自由度 $k-1$ のカイ二乗分布を表す．

自由度 $k-1$ のカイ二乗分布の上側 $100\alpha\%$ 点を $\chi_{k-1}^2(\alpha)$ とおく．このとき，次の定理 1 を導くことができる．

定理 1 $\{$ 帰無仮説 $H_{\boldsymbol{c}}$ vs. 対立仮説 $H_{\boldsymbol{c}}^A \mid \boldsymbol{c} \in \mathcal{C}^k\}$ に対する水準 α の漸近的な多重比較検定法は以下で与えられる．

(1) $T_{\boldsymbol{c}} > \chi_{k-1}^2(\alpha)$ ならば，帰無仮説 $H_{\boldsymbol{c}}$ を棄却し，対立仮説 $H_{\boldsymbol{c}}^A$を受け入れ，$\sum\limits_{i=1}^{k} c_i \log \mu_i \neq 0$ と判定する．

(2) $T_{\boldsymbol{c}} < \chi_{k-1}^2(\alpha)$ ならば，帰無仮説 $H_{\boldsymbol{c}}$ を棄却しない．

5 漸近的な同時信頼区間

(4) と同様に，

$$T_{\boldsymbol{c}}^*(\boldsymbol{\mu}) \leqq T_0(\boldsymbol{\mu})$$

が成り立ち，$T_0(\boldsymbol{\mu})$ は漸近的に自由度 $k-1$ のカイ二乗分布に従う．ここで，

$$\lim_{n\to\infty} P\left(\text{任意の } \boldsymbol{c} \in \mathcal{C}^k \text{ に対して } T_{\boldsymbol{c}}^*(\boldsymbol{\mu}) < \chi_{k-1}^2(\alpha)\right) \geqq \lim_{n\to\infty} P\left(T_0(\boldsymbol{\mu}) < \chi_{k-1}^2(\alpha)\right) = 1 - \alpha$$

が示される．これにより，次の定理 2 を導くことができる．

定理 2 $\boldsymbol{c} \in \mathcal{C}^k$ に対して，$\sum\limits_{i=1}^{k} c_i \log \mu_i$ についての信頼係数 $1-\alpha$ の漸近的な同時信頼区間は，

$$\sum_{i=1}^{k} c_i \log \hat{\mu}_i - \sqrt{\chi_{k-1}^2(\alpha) \sum_{i=1}^{k} \frac{c_i^2}{n_i}} < \sum_{i=1}^{k} c_i \log \mu_i < \sum_{i=1}^{k} c_i \log \hat{\mu}_i + \sqrt{\chi_{k-1}^2(\alpha) \sum_{i=1}^{k} \frac{c_i^2}{n_i}}$$

で与えられる．

6 C 言語プログラムによるデータ解析

6.1 プログラムの解説

文献 [4] を参考にして，多重比較検定と同時信頼区間の結果を求める C 言語プログラムを作成した．プログラムの詳細な内容は，卒業論文の第 6 章に載せている．

6.2 データの内容

2022 年 3 月 16 日に発生した福島県沖地震で，特に大きな揺れ (震度 5 弱以上) を観測した地域である，東北地方，茨城県，栃木県，新潟県を震源地とする，マグニチュード 3 以上の地震の発生間隔をデータとして解析を行った．

表 2 各群の内容

群	サイズ	期間
第 1 群	86	2021 年 10 月〜12 月
第 2 群	73	2022 年 3 月 16 日〜3 月 22 日
第 3 群	30	2022 年 3 月 23 日〜3 月 29 日
第 4 群	131	2022 年 4 月〜6 月

6.3 実行結果と考察

$\alpha = 0.05$ として，すべての対比較について C 言語プログラムにより解析を行った結果，すべての帰無仮説で棄却された．そして同時信頼区間の結果より，地震の発生間隔について次の不等式を得た．

$$\mu_2 < \mu_3 < \mu_4 < \mu_1$$

$\Leftrightarrow$ 2022 年 3/16〜3/22 $<$ 2022 年 3/23〜3/29 $<$ 2022 年 4 月〜6 月 $<$ 2021 年 10 月〜12 月

地震の発生間隔が短いということは，地震が多く起きていると考えることができる．ゆえに，福島県沖地震で特に大きな揺れを観測した地域を震源地とする地震は，2022 年 3/16〜3/22 が最も多く，福島県沖地震が発生してから時間が経つにつれて減少していることが分かる．また，2022 年 4 月〜6 月の方が，2021 年 10 月〜12 月よりも多くの地震が発生していることが分かる．

7 おわりに

本研究では，指数分布モデルにおける対数対比の多重比較法について考察した．そして，C 言語プログラムを用いて実際のデータを解析することで，より理解を深めることができた．

参考文献

[1] 髙田秀真 · 田中友規 :『多群指数モデルにおける一様性の仮説検定法』. 2018.

[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』. 2015.

[5] 日本気象協会 : https://www.jwa.or.jp/
2022 年 12 月 1 日閲覧