多群ポアソンモデルにおける対照群との統計的推測

2018SS062 高橋凌太朗

指導教員：白石高章

1　はじめに

ポアソン少数の法則により, 稀に起こる現象の回数はポアソン分布に従う. 4 年時のゼミでポアソンモデルについて学習し, 2 標本までの推測法を学習した. 本研究では多標本の推測法について論じる. まずは 2 標本モデルにおける推測法について説明し, 次に k 標本モデルにおける推測法とボンフェローニの方法による多重比較検定について述べる. データ解析では, 交通死亡事故データを使用し, ボンフェローニの方法について考察していく.

2　2 標本ポアソンモデルの推測法

$X_1, \cdots, X_{n_1}$ をポアソン分布 $\mathcal{P}_o(\mu_1)$ からの無作為標本とし $Y_1, \cdots, Y_{n_2}$ をポアソン分布 $\mathcal{P}_o(\mu_2)$ からの無作為標本とする. さらに, $(X_1, \cdots, X_{n_1})$ と $(Y_1, \cdots, Y_{n_2})$ は互いに独立とする.
$W_1 \equiv X_1 + \cdots + X_{n_1}$, $W_2 \equiv Y_1 + \cdots + Y_{n_2}$ とおく.
このとき, μ_i の点推定量は,

$$\hat{\mu}_i \equiv \frac{W_i}{n_i} \qquad (i = 1, 2)$$

で与えられる. $n \equiv n_1 + n_2$ とおき, $0 < \lim_{n\to\infty} \frac{n_1}{n} = \lambda < 1$ を仮定する. このとき, 中心極限定理 (白石 [1]) を用いて

$$\sqrt{n}(\hat{\mu}_1 - \mu_1) \xrightarrow{\mathcal{L}} N\left(0, \frac{\mu_1}{\lambda}\right)$$

$$\sqrt{n}(\hat{\mu}_2 - \mu_2) \xrightarrow{\mathcal{L}} N\left(0, \frac{\mu_2}{1-\lambda}\right)$$

が成り立つ. また, 白石 [1] の命題 7.8 を用いて

$$\sqrt{n}\{g(\hat{\mu}_1) - g(\mu_1)\} \xrightarrow{\mathcal{L}} Y_1 \sim N\left(0, \{g'(\mu_1)\}^2 \cdot \frac{\mu_1}{\lambda}\right)$$

$$\sqrt{n}\{g(\hat{\mu}_2) - g(\mu_2)\} \xrightarrow{\mathcal{L}} Y_2 \sim N\left(0, \{g'(\mu_2)\}^2 \cdot \frac{\mu_2}{1-\lambda}\right)$$

を得る.
(1)　$g(x) = \log x$ のとき

$$\sqrt{n}\{\log(\hat{\mu}_1) - \log(\mu_1)\} \xrightarrow{\mathcal{L}} Y_1 \sim N\left(0, \frac{1}{\lambda\mu_1}\right)$$

$$\sqrt{n}\{\log(\hat{\mu}_2) - \log(\mu_2)\} \xrightarrow{\mathcal{L}} Y_2 \sim N\left(0, \frac{1}{(1-\lambda)\mu_2}\right)$$

$$\sqrt{n}[\{\log(\hat{\mu}_1) - \log(\mu_1)\} - \{\log(\hat{\mu}_2) - \log(\mu_2)\}]$$
$$\xrightarrow{\mathcal{L}} Y_1 - Y_2 \sim N\left(0, \frac{1}{\lambda\mu_1} + \frac{1}{(1-\lambda)\mu_2}\right)$$

$$T(\mu) \equiv \sqrt{n}\frac{[\{\log(\hat{\mu}_1) - \log(\mu_1)\} - \log(\hat{\mu}_2) - \log(\mu_2)]}{\tilde{\sigma}}$$
$$\xrightarrow{\mathcal{L}} N(0, 1)$$

となる. ただし, $\tilde{\sigma} \equiv \sqrt{\frac{n}{\hat{\mu}_1 n_1} + \frac{n}{\hat{\mu}_2 n_2}}$ とする.

(2) $g(x) = \sqrt{x}$ のとき.

$$\sqrt{n}(\sqrt{\hat{\mu}_1} - \sqrt{\mu_1}) \xrightarrow{\mathcal{L}} Y_1 \sim N\left(0, \frac{1}{4\lambda}\right)$$

$$\sqrt{n}(\sqrt{\hat{\mu}_2} - \sqrt{\mu_2}) \xrightarrow{\mathcal{L}} Y_2 \sim N\left(0, \frac{1}{4(1-\lambda)}\right)$$

$$\sqrt{n}[\{g(\hat{\mu}_1) - g(\mu_1)\} - \{g(\hat{\mu}_2) - g(\mu_2)\}]$$
$$\xrightarrow{\mathcal{L}} Y_1 - Y_2 \sim N\left(0, \{g'(\mu_1)\}^2 \cdot \frac{\mu_1}{\lambda} + \{g'(\mu_2)\}^2 \cdot \frac{\mu_2}{1-\lambda}\right)$$

$$T(\mu) \equiv \sqrt{n}\frac{[\{g(\hat{\mu}_1) - g(\mu_1)\} - \{g(\hat{\mu}_2) - g(\mu_2)\}]}{\tilde{\sigma}} \xrightarrow{\mathcal{L}} N(0, 1)$$

となる. ただし, $\tilde{\sigma} \equiv \sqrt{\{g'(\hat{\mu}_1)\}^2 \cdot \frac{n\hat{\mu}_1}{n_1} + \{g'(\hat{\mu}_2)\}^2 \cdot \frac{n\hat{\mu}_2}{n_2}}$ とする.

$A \equiv \{|T(\mu)| < z(\alpha/2)\}$ とすると,

$$\lim_{n\to\infty} P(A^c) = \lim_{n\to\infty} 2P(T(\mu) \geq z(\alpha/2)) = \alpha$$

が成り立つ. したがって,

$$\left|\sqrt{n}\frac{\{g(\hat{\mu}_1) - g(\mu_1)\} - \{g(\hat{\mu}_2) - g(\mu_2)\}}{\tilde{\sigma}}\right| < z(\alpha/2)$$

より, $g(\mu_1) - g(\mu_2)$ に対する信頼係数 $1-\alpha$ の漸近的な信頼区間は

$$g(\hat{\mu}_1) - g(\hat{\mu}_2) - z(\alpha/2)\frac{\tilde{\sigma}}{\sqrt{n}} < g(\mu_1) - g(\mu_2)$$
$$< g(\hat{\mu}_1) - g(\hat{\mu}_2) + z(\alpha/2)\frac{\tilde{\sigma}}{\sqrt{n}}$$

で与えられる.

3　k 標本モデルにおける対照群との相違に関する多重比較

水準 A_i における観測値を $(X_{i1}, \cdots, X_{in_i})$ は第 i 標本または第 i 群と呼ばれている. X_{ij} は平均 μ_i のポアソン分布 $\mathcal{P}_o(\mu_i)$ に従うものとする. さらに, 全ての X_{ij} は互いに独立であるとする. すなわち,
$X_{ij} \sim \mathcal{P}_o(\mu_i)$ $(i = 1, \cdots, k, j = 1, \cdots, n_i)$

である. 第 k 標本を対照標本, 第 1 標本から第 $k-1$ 標本を処理標本とし, 下の表 1 のモデルについて考察する.

表 1　k 標本ポアソンモデル

水準	標本	データ
処理 1	第一標本	$X_{11},\cdots,X_{1n_1}$
処理 2	第二標本	$X_{21},\cdots,X_{2n_2}$
⋮	⋮	⋮
処理 k-1	第 k-1 標本	$X_{k-11},\cdots,X_{k-1n_{k-1}}$
対照	第 k 標本	$X_{k1},\cdots,X_{kn_k}$

第 k 標本の対照標本と第 i 標本の処理標本を比較する.

$$T \equiv \frac{\sqrt{n_i+n_k}\{g(\hat{\mu}_i)-g(\hat{\mu}_k)\}}{\tilde{\sigma}_{in}}$$

とする. ただし,

$$\bar{X}_i \equiv \hat{\mu}_i \equiv \frac{X_{i1}+X_{i2}+\cdots+X_{in_i}}{n_i}$$

$$\tilde{\sigma}_{in} \equiv \sqrt{\frac{n_i+n_k}{\hat{\mu}_i n_i}+\frac{n_i+n_k}{\hat{\mu}_k n_k}}$$

$$T_i(\boldsymbol{\mu}) \equiv \sqrt{n_i+n_k}\frac{\{g(\hat{\mu}_i)-g(\mu_i)\}-\{g(\hat{\mu}_k)-g(\mu_k)\}}{\tilde{\sigma}_{in}}$$

$$\boldsymbol{\mu} = (\mu_1,\mu_2,\cdots,\mu_k)$$

とおく.$1 \le i \le k-1$ を満たすすべての i に対して $g(\mu_i)-g(\mu_k)$ の区間推定について興味があるものとする. 定数 $\alpha(0<\alpha<1)$ を決める.
$B_i \equiv \{|T_i(\boldsymbol{\mu})| < z(\alpha/2(k-1))\}$ とおく.

$$\begin{aligned}\lim_{n\to\infty} P(B_i^c) &= \lim_{n\to\infty} P\left(|T_i(\boldsymbol{\mu})| \ge z\left(\frac{\alpha}{2(k-1)}\right)\right)\\ &= \lim_{n\to\infty} 2P\left(T_i(\boldsymbol{\mu}) \ge z\left(\frac{\alpha}{2(k-1)}\right)\right)\\ &= \frac{\alpha}{k-1} \qquad (1)\end{aligned}$$

が成り立つ. (1) とボンフェローニの不等式より

$$\lim_{n\to\infty} P(i=1,\cdots,k-1 \text{ に対して } |T_i(\boldsymbol{\mu})| < z(\alpha/2(k-1)))$$

$$= \lim_{n\to\infty} P\left(\bigcap_{i=1}^{k-1} B_i\right)$$

$$= 1-\lim_{n\to\infty} P\left(\bigcup_{i=1}^{k-1} B_i^c\right) \ge 1-\lim_{n\to\infty}\sum_{i=1}^{k-1} P(B_i^c) = 1-\alpha$$

この不等式により $\{g(\mu_i)-g(\mu_k)|1 \le i \le k-1\}$ に対する信頼係数 $1-\alpha$ の漸近的な同時信頼区間は

$$g(\hat{\mu}_i)-g(\hat{\mu}_k)-z\left(\frac{\alpha}{2(k-1)}\right)\frac{\tilde{\sigma}_{in}}{\sqrt{n_i+n_k}} < g(\mu_i)-g(\mu_k)$$

$$< g(\hat{\mu}_i)-g(\hat{\mu}_k)+z\left(\frac{\alpha}{2(k-1)}\right)\frac{\tilde{\sigma}_{in}}{\sqrt{n_i+n_k}}(1 \le i \le k-1)$$

で与えられる.

4　C 言語プログラムによるデータ解析

4.1　プログラムの解説

これまでに述べたポアソン比による検定結果を水準 $\alpha = 0.05$ で C 言語プログラムを作成した.

4.2　データの内容

データは, 警察庁の道路の交通に関する統計 [4] より, 2021 年上半期における交通死亡事故のうち, 10 都道府県を標本とし, 愛知県を対照群とする.

表 2　10 都道府県の上半期交通事故件数と死者数

都道府県	交通事故件数	交通事故死者数
北海道	3859	52
東京都	13002	57
埼玉県	7929	61
神奈川県	10559	52
千葉県	6466	59
大阪府	11840	71
兵庫県	8102	49
広島県	2159	37
福岡県	9637	38
愛知県	11578	51

4.3　実行結果

北海道, 埼玉県, 千葉県, 広島県を棄却することができた. また, 対照群である愛知県と比較した際に信頼区間が 1.0 より小さい標本が無かったことから, 9 都道府県は愛知県より交通事故による死亡率が高いという結論を得た.

5　終わりに

本研究では, k 標本モデルにおける推測法について論じてきた. また, k 標本ポアソンモデルにおけるボンフェローニの方法による多重比較検定法についても考察してきた. 実際にデータを用いて C 言語プログラムによるデータ解析も行った. 本研究を通じて多重比較検定についての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, 2019 年

[2] 井上博貴, 「ポアソンモデルにおける対照群との平均比の統計的解析」, 2020 年

[3] 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 年

[4] 警察庁交通局「令和 3 年上半期における交通死亡事故の発生状況及び道路交通法違反取締り状況等について」2021/12/06 閲覧