ポアソンモデルにおける傾向性の Cochran-Armitage 型検定

2010SE024 福地　淳史
2011SE183 名田　侑樹
指導教員：白石高章

1　はじめに

ポアソン分布に従う観測値からなるデータは, 地震の回数, 交通事故の件数など, 身の回りに多く存在する. そこで, 我々は医療事故の件数もポアソン分布に従うことを知り, 主に医療分野において活躍する Cochran-Armitage 検定に興味を持ち, 本研究を行うことを決めた. [1] を参照し比率モデルをポアソンモデルに変えた Cochran-Armitage 型検定について考察する.

2　標本平均を使った Cochran-Armitage 型検定

ある定量的, あるいは順序尺度変数 X の a 個の値 $x_1, \cdots, x_a$ において, ポアソン分布に従う応答変数 Y が観測されているとする. ここで x_i において ,n_i 個の $Y_{i1}, \cdots, Y_{in_i}$ を観測する $(i = 1, \cdots, a)$. $\{Y_{ij} | j = 1, \cdots, n_i; i = 1, \cdots, a\}$ は互いに独立とし,

$$E(Y_{ij}) = \mu(x_i)$$

とする. $\mu(x)$ は $X = x_i$ で $\mu(x_i)$ をとる広義の単調関数とする. すなわち,

$$\mu(x_1) \leqq \cdots \leqq \mu(x_a) \text{ または } \mu(x_1) \geqq \cdots \geqq \mu(x_a)$$

である. x_i に評点 $s(x_i)$ を与え, ここに,

$$s(x_1) < \cdots < s(x_a)$$

とする. さらに W_i を,

$$W_i = \sum_{j=1}^{n_i} Y_{ij}$$

とおく. $\mu(x_i)$ を μ_i, $s(x_i)$ を s_i とおくと, このとき, $Y_{ij} \sim P_o(\mu_i)$ となる. 表 1 は [1] を参照.

表 1　母平均モデルと観測値

要因	$x_1 \ \cdots \ x_i \ \cdots \ x_a$
事象発生数	$W_1 \ \cdots \ W_i \ \cdots \ W_a$
標本サイズ	$n_1 \ \cdots \ n_i \ \cdots \ n_a$
評点	$s_1 \ \cdots \ s_i \ \cdots \ s_a$
母平均	$\mu_1 \ \cdots \ \mu_i \ \cdots \ \mu_a$

母平均 μ_i の推定値は, $\hat{\mu}_i$ は,

$$\hat{\mu}_i = \frac{W_i}{n_i}$$

である. また, n, $\bar{\mu}$, $\bar{\hat{\mu}}$, $\bar{s}$ を,

$$n \equiv \sum_{i=1}^{a} n_i \ , \ \bar{\mu} \equiv \sum_{i=1}^{a} \frac{n_i \mu_i}{n}$$
$$\bar{\hat{\mu}} \equiv \sum_{i=1}^{a} \frac{n_i \hat{\mu}_i}{n} = \sum_{i=1}^{a} \frac{W_i}{n} \ , \ \bar{s} \equiv \sum_{i=1}^{a} \frac{n_i s_i}{n}$$

命題 1. 確率変数 Y の評点 s 上への単回帰係数の推定値を $\hat{b}$ とすると,

$$\hat{b} = \frac{\sum_{i=1}^{a} n_i (s_i - \bar{s})(\hat{\mu}_i - \bar{\hat{\mu}})}{\sum_{i=1}^{a} n_i (s_i - \bar{s})^2}$$

で与えられる.

回帰係数の推定値 $\hat{b}$ の期待値 $E(\hat{b})$ は,

$$\begin{aligned}
E(\hat{b}) &= E\left(\frac{\sum_{i=1}^{a} n_i (s_i - \bar{s})(\hat{\mu}_i - \bar{\hat{\mu}})}{\sum_{i=1}^{a} n_i (s_i - \bar{s})^2} \right) \\
&= \frac{\sum_{i=1}^{a} E\left(n_i (s_i - \bar{s})(\hat{\mu}_i - \bar{\hat{\mu}}) \right)}{\sum_{i=1}^{a} n_i (s_i - \bar{s})^2} \\
&= \frac{\sum_{i=1}^{a} n_i (s_i - \bar{s})(\mu_i - \bar{\mu})}{\sum_{i=1}^{a} n_i (s_i - \bar{s})^2}
\end{aligned}$$

となる. また,

$$\begin{aligned}
(\hat{b}\text{の分子}) &= \sum_{i=1}^{a} n_i (s_i - \bar{s})(\hat{\mu}_i - \bar{\hat{\mu}}) \\
&= \sum_{i=1}^{a} (n_i s_i \hat{\mu}_i - n_i \bar{s} \hat{\mu}_i + n_i \bar{\hat{\mu}} \bar{s} - n_i \bar{\hat{\mu}} s_i)
\end{aligned}$$

である. ここで,

$$\begin{aligned}
\sum_{i=1}^{a} (n_i \bar{s} \bar{\hat{\mu}} - n_i \bar{\hat{\mu}} s_i) &= \bar{\hat{\mu}} \bar{s} \sum_{i=1}^{a} n_i - \bar{\hat{\mu}} \sum_{i=1}^{a} n_i s_i \\
&= \bar{\hat{\mu}} \bar{s} n - \bar{\hat{\mu}} \bar{s} n = 0
\end{aligned}$$

である. したがって,

$$(\hat{b}\text{の分子}) = \sum_{i=1}^{a}(n_i s_i \hat{\mu}_i - n_i \bar{s}\hat{\mu}_i) = \sum_{i=1}^{a} n_i \hat{\mu}_i (s_i - \bar{s})$$

である. 回帰係数の推定値 $\hat{b}$ の分散は,

$$V(\hat{b}) = V\left(\frac{\sum_{i=1}^{a} n_i \hat{\mu}_i (s_i - \bar{s})}{\sum_{i=1}^{a} n_i (s_i - \bar{s})^2}\right)$$

$$= \frac{V\left(\sum_{i=1}^{a} n_i \hat{\mu}_i (s_i - \bar{s})\right)}{\left(\sum_{i=1}^{a} n_i (s_i - \bar{s})^2\right)^2} \tag{2.1}$$

ここで,

$$\begin{aligned}((2.1)\text{ の分子}) &= V\left(\sum_{i=1}^{a} n_i \hat{\mu}_i (s_i - \bar{s})\right) \\ &= \sum_{i=1}^{a} V\left(n_i \hat{\mu}_i (s_i - \bar{s})\right) \\ &= \sum_{i=1}^{a} n_i^2 (s_i - \bar{s})^2 V(\hat{\mu}_i)\end{aligned}$$

また,

$$V(\hat{\mu}_i) = V\left(\frac{W_i}{n_i}\right) = \frac{V(W_i)}{n^2{}_i} = \frac{n_i \mu_i}{n^2{}_i} = \frac{\mu_i}{n_i}$$

となるので,

$$((2.1)\text{ の分子}) = \sum_{i=1}^{a} n_i \mu_i (s_i - \bar{s})^2$$

である. したがって,

$$V(\hat{b}) = \frac{\sum_{i=1}^{a} n_i \mu_i (s_i - \bar{s})^2}{\left(\sum_{i=1}^{a} n_i (s_i - \bar{s})^2\right)^2}$$

である. 次に,

帰無仮説 $H_0 : \mu_1 = \cdots = \mu_a$

対立仮説 $H_1 : \mu_1 \leqq \cdots \leqq \mu_a$ または $\mu_1 \geqq \cdots \geqq \mu_a$(少なくとも 1 つの不等号は等号が成り立たないものとする)

とおく. また, H_0 の下で $\mu_1 = \cdots = \mu_a = \mu_0$ とする. H_0 の下での $\hat{b}$ の分散 $V_0(\hat{b})$ と $V_0(\hat{b})$ の推定値 $\widehat{V_0(\hat{b})}$ は,

$$V_0(\hat{b}) = \frac{\sum_{i=1}^{a} n_i \mu_0 (s_i - \bar{s})^2}{\left(\sum_{i=1}^{a} n_i (s_i - \bar{s})^2\right)^2} = \frac{\mu_0}{\sum_{i=1}^{a} n_i (s_i - \bar{s})^2}$$

$$\widehat{V_0(\hat{b})} = \frac{\bar{\hat{\mu}}}{\sum_{i=1}^{a} n_i (s_i - \bar{s})^2}$$

である. また, H_0 の下での $\hat{b}$ の期待値 $E_0(\hat{b})$ は,

$$E_0(\hat{b}) = \frac{\sum_{i=1}^{a} n_i (s_i - \bar{s})(\mu_0 - \bar{\mu})}{\left(\sum_{i=1}^{a} n_i (s_i - \bar{s})^2\right)^2} = 0$$

である. さらに T_b を,

$$T_b = \frac{\hat{b} - E_0(\hat{b})}{\sqrt{V_0(\hat{b})}}$$

とおく. 次の (条件 1) を仮定する.

(条件 1) $\quad 0 < \lambda_i \equiv \lim_{n\to\infty} \frac{n_i}{n} < 1$

H_0 の下で $n\to\infty$ として,

$$\begin{aligned}T_b &= \frac{\sum_{i=1}^{a} n_i (s_i - \bar{s})(\hat{\mu}_i - \bar{\mu})}{\sqrt{\mu_0 \sum_{i=1}^{a} n_i (s_i - \bar{s})^2}} = \frac{\sum_{i=1}^{a} n_i (s_i - \bar{s})(\hat{\mu}_i - \mu_0)}{\sqrt{\mu_0 \sum_{i=1}^{a} n_i (s_i - \bar{s})^2}} \\ &= \frac{\sum_{i=1}^{a} \frac{n_i}{n}(s_i - \bar{s})\sqrt{n}(\hat{\mu}_i - \mu_0)}{\sqrt{\mu_0 \sum_{i=1}^{a} \frac{n_i}{n}(s_i - \bar{s})^2}} \\ &\xrightarrow{\mathcal{L}} \frac{\sum_{i=1}^{a} \lambda_i (s_i - \bar{s})\sqrt{\frac{\mu_i}{\lambda_i}} Z_i}{\sqrt{\mu_0 \sum_{i=1}^{a} \lambda_i (s_i - \bar{s})^2}} \\ &\sim N\left(0\,,\ \frac{\sum_{i=1}^{a} \lambda_i^2 (s_i - \bar{s})^2 \frac{\mu_0}{\lambda_i} V(Z_i)}{\mu_0 \sum_{i=1}^{a} \lambda_i (s_i - \bar{s})^2}\right) = N(0,1)\end{aligned} \tag{2.2}$$

である. ただし, $A_n \xrightarrow{\mathcal{L}} A$ は A_n が A に分布収束することを示し, $Z_i \sim N(0,1)$ である. さらに, v_0, $\hat{v}_0$ を次のようにおく.

$$v_0 \equiv V_0(\hat{b}) = \frac{\mu_0}{\sum_{i=1}^{a} n_i (s_i - \bar{s})^2}$$

$$\hat{v}_0 \equiv \widehat{V_0(\hat{b})} = \frac{\bar{\hat{\mu}}}{\sum_{i=1}^{a} n_i (s_i - \bar{s})^2} \xrightarrow{P} v_0 \tag{2.3}$$

ただし, $A_n \xrightarrow{P} A$ は A_n が A に確率収束することを示している. (2.2), (2.3) とスラツキーの定理より,

$$\frac{\sqrt{v_0}}{\sqrt{\hat{v}_0}} \frac{\hat{b}}{\sqrt{v_0}} = \frac{\hat{b}}{\sqrt{\hat{v}_0}} \xrightarrow{\mathcal{L}} \frac{\sqrt{v_0}}{\sqrt{\hat{v}_0}} N(0,1) = N(0,1)$$

である. したがって白石 [2] の定理 3.17 より,

$$\frac{\hat{b}^2}{\hat{v}_0} \xrightarrow{\mathcal{L}} \chi_1^2$$

となる. このとき, 水準 α の検定は, $\hat{b}^2/\hat{v}_0 > \chi_1^2(\alpha)$ のとき H_0 を棄却し, $\hat{b}^2/\hat{v}_0 < \chi_1^2(\alpha)$ のとき H_0 を棄却しない. ただし, $\chi_1^2(\alpha)$ は, 自由度 1 の χ^2 分布の上側 $100\alpha\%$ 点である.

3 分散安定化変換を使った Cochran-Armitage 型検定

白石 [2] の (8.17) より,

$$\sqrt{n_i}(\hat{\mu}_i - \mu_i) \xrightarrow{\mathcal{L}} Y \sim N(0, \mu_i)$$

が成り立ち, また白石 [2] の命題 7.8 より,

$$T_n = \hat{\mu}_i, \quad \theta = \mu_i, \quad h(\mu_i) = \mu_i, \quad g(\hat{\mu}_i) = \sqrt{\hat{\mu}_i} = \hat{\sigma}_i,$$

$$g(\mu_i) = \sqrt{\mu_i} = \sigma_i, \quad g'(\mu_i) = \frac{1}{2\sqrt{\mu_i}}$$

と適応させると,

$$\sqrt{n_i}(\hat{\sigma}_i - \sigma_i) \xrightarrow{\mathcal{L}} N\left(0, \left\{\frac{1}{2\sqrt{\mu_i}}\right\}^2 \cdot \mu_i\right) = N\left(0, \frac{1}{4}\right)$$

となる .

$\sqrt{\dfrac{n}{n_i}} \xrightarrow{P} \sqrt{\dfrac{1}{\lambda_i}}$ となることとスラツキーの定理より,

$$\begin{aligned} & \sqrt{n_i}(\hat{\sigma}_i - \sigma_i) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{4}\right) \\ \Longleftrightarrow\ & 2\sqrt{n_i}(\hat{\sigma}_i - \sigma_i) \xrightarrow{\mathcal{L}} N(0,1) \\ \Longleftrightarrow\ & \sqrt{\frac{n}{n_i}} 2\sqrt{n_i}(\hat{\sigma}_i - \sigma_i) \xrightarrow{\mathcal{L}} \sqrt{\frac{1}{\lambda_i}} N(0,1) \\ \Longleftrightarrow\ & 2\sqrt{n}(\hat{\sigma}_i - \sigma_i) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{\lambda_i}\right) \qquad (3.1) \end{aligned}$$

が成り立つ.

補題 2. (条件 1) を満たすものとする. b_s を,

$$b_s \equiv \frac{\displaystyle\sum_{i=1}^{a} \left(\frac{n_i}{n}\right)(s_i - \bar{s}) 2\sqrt{n}(\hat{\sigma}_i - \sigma_i)}{\sqrt{\displaystyle\sum_{i=1}^{a} (s_i - \bar{s})^2 \left(\frac{n_i}{n}\right)}}$$

とおくと $n \to \infty$ として, $b_s \xrightarrow{\mathcal{L}} N(0,1)$ である.

(証明) (3.1) より,

$$\begin{aligned} (b_s\text{の分子}) &= \sum_{i=1}^{a} \left(\frac{n_i}{n}\right)(s_i - \bar{s}) 2\sqrt{n}(\hat{\sigma}_i - \sigma_i) \\ &\xrightarrow{\mathcal{L}} \sum_{i=1}^{a} \lambda_i (s_i - \bar{s}) \frac{Z_i}{\sqrt{\lambda_i}} \\ &= \sum_{i=1}^{a} \sqrt{\lambda_i}(s_i - \bar{s}) Z_i \end{aligned}$$

となる. ただし,$Z_i \sim N(0,1)$ である.
したがって, $E(Z_i) = 0,\ V(Z_i) = 1$ となる. また,

$$(b_s\text{の分母}) = \sqrt{\sum_{i=1}^{a} (s_i - \bar{s})^2 \left(\frac{n_i}{n}\right)} \xrightarrow{P} \sqrt{\sum_{i=1}^{a} (s_i - \bar{s})^2 \lambda_i}$$

となるので, スラツキーの定理より,

$$\begin{aligned} b_s &\xrightarrow{\mathcal{L}} \frac{\displaystyle\sum_{i=1}^{a} \sqrt{\lambda_i}(s_i - \bar{s}) Z_i}{\sqrt{\displaystyle\sum_{i=1}^{a} (s_i - \bar{s})^2 \lambda_i}} \\ &\sim N\left(0\ ,\ \frac{\displaystyle\sum_{i=1}^{a} \left\{\sqrt{\lambda_i}(s_i - \bar{s})\right\}^2 V(Z_i)}{\left\{\sqrt{\displaystyle\sum_{i=1}^{a} (s_i - \bar{s})^2 \lambda_i}\right\}^2}\right) = N(0,1) \end{aligned}$$

が成り立つ. □

ここで $\bar{\hat{\sigma}}$ を,

$$\bar{\hat{\sigma}} \equiv \sum_{i=1}^{a} \left(\frac{n_i}{n}\right) \hat{\sigma}_i$$

とおく.

定理 3. (条件 1) を満たすものとする. このとき $n \to \infty$ とすると, H_0 の下で,

$$\hat{b}_s \equiv \frac{\displaystyle\sum_{i=1}^{a} \left(\frac{n_i}{n}\right)(s_i - \bar{s}) 2\sqrt{n}\left(\hat{\sigma}_i - \bar{\hat{\sigma}}\right)}{\sqrt{\displaystyle\sum_{i=1}^{a} (s_i - \bar{s}) \left(\frac{n_i}{n}\right)}} \xrightarrow{\mathcal{L}} N(0,1)$$

が成り立つ.

(証明) 補題 2 より, H_0 の下で, $\hat{b}_s - b_s \xrightarrow{P} 0$ を示せばよい.

$$\begin{aligned} & \left(\hat{b}_s - b_s\text{の分子}\right) \\ &= \sum_{i=1}^{a} \left(\frac{n_i}{n}\right)(s_i - \bar{s}) 2\sqrt{n}\left(\hat{\sigma}_i - \bar{\hat{\sigma}}\right) \\ &\quad - \sum_{i=1}^{a} \left(\frac{n_i}{n}\right)(s_i - \bar{s}) 2\sqrt{n}(\hat{\sigma}_i - \sigma_1) \\ &= \sum_{i=1}^{a} \left(\frac{n_i}{n}\right)(s_i - \bar{s}) 2\sqrt{n}\hat{\sigma}_i - \sum_{i=1}^{a} \left(\frac{n_i}{n}\right)(s_i - \bar{s}) 2\sqrt{n}\bar{\hat{\sigma}} \end{aligned}$$

$$-\sum_{i=1}^{a}\left(\frac{n_i}{n}\right)(s_i-\bar{s})2\sqrt{n}\hat{\sigma}_i \ - \ \sum_{i=1}^{a}\left(\frac{n_i}{n}\right)(s_i-\bar{s})2\sqrt{n}\sigma_1$$

$$=-\frac{2\bar{\hat{\sigma}}}{\sqrt{n}}\left\{\sum_{i=1}^{a}n_i(s_i-\bar{s})\right\} \ - \ \frac{2\sigma_1}{\sqrt{n}}\left\{\sum_{i=1}^{a}n_i(s_i-\bar{s})\right\}$$

ここで,

$$\sum_{i=1}^{a}n_i(s_i-\bar{s})=\sum_{i=1}^{a}n_is_i \ - \ \sum_{i=1}^{a}n_i\bar{s}$$
$$=n\bar{s} \ - \ \bar{s}\sum_{i=1}^{a}n_i \ =n\bar{s} \ - \ n\bar{s} \ = \ 0$$

である．よって，$\hat{b}_s-b_s=0$ となるので, $\hat{b}_s-b_s\xrightarrow{P}0$ である． □

白石 [2] の定理 3.17 より,

$$\hat{b}_s^2\xrightarrow{\mathcal{L}}\chi_1^2$$

となる. このとき, 水準 α の検定は, $\hat{b}_s^2>\chi_1^2(\alpha)$ のとき H_0 を棄却し, $\hat{b}_s^2<\chi_1^2(\alpha)$ のとき H_0 を棄却しない.

4 C言語プログラム解説

C 言語により,Cochran-Armitage 型検定による検定結果及び, ポアソン分布に従う変数を用いた検定結果を作成した. 本研究で作成したプログラムの main プログラムは,

```
int main(void){
input();
keisan1();
keisan2();
kentei1();
kentei2();
return (0);
}
```

である.

1.input 関数の中で, 順序尺度変数, 有意水準, 標本サイズを入力する. ただし, $s_i=i$ とする.

2.keisan1 関数の中で, W_i(W[i]) , $\hat{\mu}_i$(mh[i]) , n(m) , $\bar{\hat{\mu}}$(mhb) , $\bar{s}$(sb) , s_i(s[i]) $\bar{\hat{\sigma}}$(shb) , $\hat{\sigma}_i$(sh[i]) の値を計算する.

3.keisan2 関数の中で, $\hat{b}$(bh) と $\hat{V}_0$(v0h) , $\hat{b}^2/\hat{v}_0$ (t) , $\chi_1^2(\alpha)$(XN1) の値を計算する.

4.kentei1 関数の中で, 標本平均を使った Cochran-Armitage 型検定を行う.

5.kentei2 関数の中で, 分散安定化変換を使った Cochran-Armitage 型を行う.

4.1 検定内容

医療関係のポアソン分布に従う変数のデータが得られなかったため, 一ヶ月に起きた地震の回数のデータ [3] をもとに検定を行った.2014 年 5 月に長野県で非常に多くの地震が発生した. 地震が発生する過去三ヶ月の 2 月から 4 月までのその月における地震の一日平均が単調増加であったため, この期間の長野県全域の震度 1 以上の地震データ [3] を入力する.

4.2 プログラム出力結果

今回はページの都合上, 有意水準 $\alpha=0.10$ のときのみを記載. ただし, 実行結果が長くなるので,W_i にそれぞれの月で起きた地震の回数を入力した場合のみの結果を記載する. 以下実行結果

順序尺度変数 X の個数 a を入力 3
有意水準 α を入力 ($0<\alpha\leqq 1$)0.10
x_1 における観測値の個数 n_1 を入力 28
x_2 における観測値の個数 n_2 を入力 31
x_3 における観測値の個数 n_3 を入力 30
順序尺度変数 x_1を入力2
順序尺度変数 x_2を入力3
順序尺度変数 x_3を入力4
観測値をそのまま入力 yes 1 , no 0
0
W_1 を入力 2
W_2 を入力 4
W_3 を入力 7
$H_0:\mu_1=\cdots=\mu_a$
$H_1:\mu_1\leqq\cdots\leqq\mu_a$ または $\mu_1\geqq\cdots\geqq\mu_a$
標本平均を使った Cochran-Armitage 型検定
H_0 を棄却しない
分散安定化変換を使った Cochran-Armitage 型検定
H_0 を棄却する

4.3 実行結果による考察

有意水準 $\alpha=0.10$ の標本平均を使った Cochran-Armitage 型検定は H_0 を棄却せず, 分散安定化変換を使った Cochran-Armitage 型検定では H_0 を棄却した. この結果より, 一般的に使われている有意水準 $\alpha=0.05$ より高い 0.10 だが,5 月に向けて一ヶ月における地震回数の平均は徐々に単調増加であるといえる.

5 おわりに

今回の研究では, ポアソン分布に従う変数を用いた検定を行い, 導くことができた. また,C 言語によって作成したプログラムによって, 同様の結果を得られた. 実際にプログラムを作成し, 現実の地震データを入力したことにより, 本研究への理解がより一層深まった.

参考文献

[1] 丹後俊朗, 小西貞則 :『医学統計学の辞典』. 朝倉書店, 東京 2012 年.

[2] 白石高章 :『統計学の基礎』. 日本評論社, 東京,2012 年.

[3] http://www.jma.go.jp/jma/index.html 気象庁ホームページ (2014 年 12 月 10 日現在)