

# 層別化された多群ポアソンモデルの統計解析法

2014SS065 岡田拓也 2013SE230 鶴飼晃伸

指導教員：白石高章

## 1 はじめに

ポアソン分布に従うデータは、単位時間における機械の故障回数や事故回数などたくさん存在する。ポアソン分布は、稀にしか起きない現象を大量に観察したデータに適用される(参考文献[1]より)。そこで、カイ二乗型検定法に興味を持ち、この検定手法について研究することにした。

本論では、層別化された多群ポアソンモデルの統計解析法について考察する。

## 2 カイ二乗型検定法

標本の観測値を  $X_{ijk}$  ( $k = 1, \dots, n_j; j = 1, \dots, J; i = 1, \dots, I$ ) とする。 $X_{ijk}$  は互いに独立と仮定する。また  $X_{ijk}$  は平均  $\mu_{ij}$  のポアソン分布  $\mathcal{P}_o(\mu_{ij})$  に従う。すなわち、 $X_{ijk} \sim \mathcal{P}_o(\mu_{ij})$  である。帰無仮説  $H_0: \mu_{i1} = \dots = \mu_{iJ}$  ( $i = 1, \dots, I$ ) vs. 対立仮説  $H_1$ : ある  $j$  と  $j'$  について、 $\mu_{ij} \neq \mu_{ij'}$  ( $i = 1, \dots, I$ ) を考える。

表1 モデル

標本	サイズ	データ	$X_{ijk}$ の分布
第1標本	$n_1$	$X_{i11}, \dots, X_{i1n_1}$	$\mathcal{P}_o(\mu_{i1})$
第2標本	$n_2$	$X_{i21}, \dots, X_{i2n_2}$	$\mathcal{P}_o(\mu_{i2})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
第 $j$ 標本	$n_j$	$X_{ij1}, \dots, X_{ijn_j}$	$\mathcal{P}_o(\mu_{ij})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
第 $J$ 標本	$n_J$	$X_{iJ1}, \dots, X_{iJn_J}$	$\mathcal{P}_o(\mu_{iJ})$

$i$  層目の標本サイズ:  $n \equiv n_1 + \dots + n_J$

次の(条件1)を仮定する。

$$(\text{条件1}) \quad \lim_{n \rightarrow \infty} \frac{n_j}{n} = \lambda_j > 0 \quad (j = 1, \dots, J)$$

命題 2.1 (文献[2]の命題8.2)

確率変数  $X, Y$  は互いに独立とする。このとき、

$$X \sim \mathcal{P}_o(\mu_1), Y \sim \mathcal{P}_o(\mu_2) \implies X + Y \sim \mathcal{P}_o(\mu_1 + \mu_2)$$

が成り立つ。□

$X_{ij\cdot} \equiv X_{ij1} + \dots + X_{ijn_j}$  で定義すると、 $X_{ijk}$  は互いに独立なので、 $X_{ij\cdot}$  の分布は、命題 2.1 より、

$$X_{ij\cdot} = X_{ij1} + \dots + X_{ijn_j} \sim \mathcal{P}_o(n_j \mu_{ij})$$

と表すことができる。

$\mu_{ij}$  の点推定量を

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

とする。

定理 2.2 (文献[2]の定理8.1)

$X \sim \mathcal{P}_o(\mu)$  ならば、平均と分散は

$$E(X) = V(X) = \mu$$

で与えられる。□

$\hat{\mu}_{ij}$  の平均と分散は定理 2.2 を用いて、

$$E(\hat{\mu}_{ij}) = \mu_{ij} \quad (j = 1, \dots, J) \quad (1)$$

$$V(\hat{\mu}_{ij}) = \frac{1}{n_j} \mu_{ij} \quad (j = 1, \dots, J) \quad (2)$$

と求めることができる。

$X_1, \dots, X_n$  を互いに独立で同一の分布に従う確率変数とし、平均と分散をそれぞれ  $\mu, \sigma^2$  とする。このとき、最初の  $n$  個の標本平均  $\bar{X}_n \equiv (X_1 + \dots + X_n)/n$  の平均と分散は、文献[2]の補題 2.25 より、

$$E(\bar{X}_n) = \mu, \quad V(\bar{X}_n) = \frac{\sigma^2}{n}$$

である。

よって、文献[2]の定理 3.27 の中心極限定理に適用する。

(1) と (2) より、

$$\begin{aligned} \frac{\hat{\mu}_{ij} - E(\hat{\mu}_{ij})}{\sqrt{V(\hat{\mu}_{ij})}} &= \frac{\hat{\mu}_{ij} - \mu_{ij}}{\sqrt{\frac{1}{n_j} \mu_{ij}}} \\ &= \frac{\sqrt{n_j}(\hat{\mu}_{ij} - \mu_{ij})}{\sqrt{\mu_{ij}}} \\ &\xrightarrow{L} Y_{ij} \sim N(0, 1) \end{aligned}$$

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

となる。

また、(条件1)のもとで、

$$\begin{aligned} \sqrt{n}(\hat{\mu}_{ij} - \mu_{ij}) &= \frac{\sqrt{n}}{\sqrt{n_j}} \cdot \sqrt{\mu_{ij}} \cdot \frac{\sqrt{n_j}(\hat{\mu}_{ij} - \mu_{ij})}{\sqrt{\mu_{ij}}} \\ &\xrightarrow{L} \frac{1}{\sqrt{\lambda_j}} \cdot \sqrt{\mu_{ij}} \cdot Y_{ij} \\ &\sim N\left(0, \frac{1}{\lambda_j} \mu_{ij}\right) \end{aligned}$$

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

ここで、文献[2]の定理 3.35 のデルタ法を用いると

となり題意は示された。□

$$\begin{aligned} 2\sqrt{n}(\hat{\sigma}_{ij} - \sigma_{ij}) &\stackrel{L}{\rightarrow} \frac{1}{\sqrt{\lambda_j}} \cdot \frac{1}{\sqrt{\mu_{ij}}} \cdot \sqrt{\mu_{ij}} Y_{ij} \\ &= \frac{1}{\sqrt{\lambda_j}} Y_{ij} \\ &\sim N\left(0, \frac{1}{\lambda_j}\right) \end{aligned}$$

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

が成り立つ。

### 3 提案する検定法

ここで統計量  $T_P$  の漸近分布を求める前に  $i$  に対しての統計量  $T_{iP}$  を求める

$$S_{ij} \equiv 2\sqrt{n}(\hat{\sigma}_{ij} - \sigma_{ij})$$

とおくと

$$S_{ij} \stackrel{L}{\rightarrow} W_{ij} \sim N\left(0, \frac{1}{\lambda_j}\right)$$

となる

**補題 3.1**  $H_0$  のもとで

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

とおくと、

$$\hat{Z}_{ij} \stackrel{L}{\rightarrow} \sqrt{\lambda_j} \left( W_{ij} - \sum_{k=1}^J \lambda_k W_{ik} \right)$$

$$(i = 1, \dots, I, j = 1, \dots, J)$$

が成り立つ。

< 証明 >

$$\begin{aligned} \hat{Z}_{ij} &= 2\sqrt{n_j} \left( \hat{\sigma}_{ij} - \sum_{k=1}^J \frac{n_k}{n} \hat{\sigma}_{ik} \right) \\ &= \frac{\sqrt{n_j}}{\sqrt{n}} \left\{ 2\sqrt{n}(\hat{\sigma}_{ij} - \sigma_{ij}) - \sum_{k=1}^J \frac{n_k}{n} 2\sqrt{n}(\hat{\sigma}_{ik} - \sigma_{ik}) \right\} \end{aligned}$$

ここで、 $S_{ij} = 2\sqrt{n}(\hat{\sigma}_{ij} - \sigma_{ij})$  より上式を書き換えると

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

となる。ここで、条件 (1) と  $S_{ij} \stackrel{L}{\rightarrow} W_{ij}$  と文献 [2] の定理 3.39 のクラメル-ウォルドのテクニックより

$$\hat{Z}_{ij} \stackrel{L}{\rightarrow} \sqrt{\lambda_j} \left( W_{ij} - \sum_{k=1}^J \lambda_k W_{ik} \right)$$

となり題意は示された。□

**定理 3.2** 統計量  $T_{iP}$  を  $T_{iP} \equiv \sum_{j=1}^J (\hat{Z}_{ij})^2$  とすると  
帰無仮説  $H_0$  において  $n \rightarrow \infty$  とすると

$$T_{iP} \stackrel{L}{\rightarrow} T \sim \chi_{J-1}^2 (i = 1, \dots, I)$$

が成立する。

< 証明 >

$T_{iP} = \sum_{k=1}^J (\hat{Z}_{ik})^2$  に補題 3.1 の条件を用いて

$$\begin{aligned} T_{iP} &= \sum_{j=1}^J \left\{ \frac{\sqrt{n_j}}{\sqrt{n}} \left( S_{ij} - \sum_{k=1}^J \frac{n_k}{n} S_{ik} \right) \right\}^2 \\ &= \sum_{j=1}^J \frac{n_j}{n} \left( S_{ij} - \sum_{k=1}^J \frac{n_k}{n} S_{ik} \right)^2 \end{aligned}$$

と書き換えることができ、これに補題 3.1 を用いることで

$$T_{iP} \stackrel{L}{\rightarrow} T \equiv \sum_{k=1}^J \lambda_i \left( W_{ik} - \sum_{j=1}^J \lambda_j W_{ij} \right)^2$$

ということが出来る。

文献 [2] の定理 3.23 より

$$T \sim \chi_{J-1}^2$$

となり題意は示された。□

次に  $\hat{Z}_{.j} \equiv \frac{1}{\sqrt{I}} \sum_{i=1}^I \hat{Z}_{ij}$  とおくと帰無仮説  $H_0$  のもとで

$$\begin{aligned} \hat{Z}_{.j} &= \frac{1}{\sqrt{I}} \sum_{i=1}^I \hat{Z}_{ij} \\ &= \frac{1}{\sqrt{I}} \sum_{i=1}^I \frac{\sqrt{n_j}}{n} \left\{ 2\sqrt{n}(\hat{\sigma}_{ij} - \sigma_{ij}) \right\} - \sum_{k=1}^J \frac{n_k}{n} 2\sqrt{n}(\hat{\sigma}_{ik} - \sigma_{ik}) \end{aligned}$$

また、 $S_{ij} = 2\sqrt{n}(\hat{\sigma}_{ij} - \sigma_{ij})$  より

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

となり、 $S_{ij} \stackrel{L}{\rightarrow} W_{ij}$  より

$$\begin{aligned} \hat{Z}_{.j} &\stackrel{L}{\rightarrow} \frac{1}{\sqrt{I}} \sum_{i=1}^I \sqrt{\lambda_j} \left( W_{ij} - \sum_{k=1}^J \lambda_k W_{ik} \right) \\ &= \sqrt{\lambda_j} \left( \frac{1}{\sqrt{I}} \sum_{j=1}^I W_{ij} - \sum_{j=1}^J \lambda_j \frac{1}{\sqrt{I}} \sum_{j=1}^I W_{ij} \right) \end{aligned}$$

となり, また  $W'_j \equiv \frac{1}{\sqrt{I}} \sum_{i=1}^I W_{ij}$  と置くと

$$\hat{Z}_{\cdot j} = \sqrt{\lambda_j} \left( W'_j - \sum_{j=1}^J \lambda_j W'_j \right)$$

と書き換えることが出来る. したがって

$$\begin{aligned} W'_j &= \sqrt{\frac{1}{I}} \sum_{i=1}^I W_{ij} \\ &\sim N \left( 0, \sum_{i=1}^I \frac{1}{I} \cdot \frac{1}{\lambda_j} \right) \\ &= N \left( 0, \frac{1}{\lambda_j} \right) \end{aligned}$$

であるといえる.

$$\text{次に } T_P \equiv \sum_{j=1}^J (\hat{Z}_{\cdot j})^2 \text{ とすると,}$$

**定理 3.3** 統計量  $T_P$  を  $T_P \equiv \sum_{j=1}^J (\hat{Z}_{\cdot j})^2$  とすると  
帰無仮説  $H_0$  において  $n \rightarrow \infty$  とすると

$$T_P \xrightarrow{L} T \sim \chi_{J-1}^2$$

が成立する.

< 証明 >

$$\begin{aligned} T_P &= \sum_{j=1}^J (\hat{Z}_{\cdot j})^2 \\ &= \sum_{j=1}^J \left[ \frac{1}{\sqrt{I}} \sum_{i=1}^I \frac{\sqrt{n_j}}{\sqrt{n}} \left\{ 2\sqrt{n} (\hat{\sigma}_{ij} - \sigma_{ij}) \right. \right. \\ &\quad \left. \left. - \sum_{k=1}^J \frac{n_k}{n} \cdot 2\sqrt{n} (\hat{\sigma}_{ik} - \sigma_{ik}) \right\} \right]^2 \end{aligned}$$

ここで,  $S_{ij} = 2\sqrt{n}(\hat{\sigma}_{ij} - \sigma_{ij})$  を変形した式に代入すると,

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

となる.

$S_{ij} \xrightarrow{L} W_{ij}$  を用いて, 上式書き換えると

$$\begin{aligned} T_P &\xrightarrow{L} \sum_{j=1}^J \left\{ \frac{1}{\sqrt{I}} \sum_{i=1}^I \sqrt{\lambda_j} \left( W_{ij} - \sum_{k=1}^J \lambda_k W_{ik} \right) \right\}^2 \\ &= \sum_{j=1}^J \lambda_j \left( \frac{1}{\sqrt{I}} \sum_{i=1}^I W_{ij} - \sum_{k=1}^J \lambda_k \frac{1}{\sqrt{I}} \sum_{i=1}^I W_{ik} \right)^2 \end{aligned}$$

となる.

また  $W'_j = \frac{1}{\sqrt{I}} \sum_{i=1}^I W_{ij}$  を用いることで,

$$\begin{aligned} T_P &\xrightarrow{L} \sum_{j=1}^J \lambda_j \left( \frac{1}{\sqrt{I}} \sum_{i=1}^I W_{ij} - \sum_{k=1}^J \lambda_k \frac{1}{\sqrt{I}} \sum_{i=1}^I W_{ik} \right)^2 \\ &= \sum_{j=1}^J \lambda_j \left( W'_j - \sum_{k=1}^J \lambda_k W'_k \right)^2 \end{aligned}$$

となる.

また上式の右辺を  $T$  とおくと, 文献 [2] の定理 3.23 より,  $T \sim \chi_{J-1}^2$  である.  $\square$

$$\begin{aligned} T_P &= \sum_{j=1}^J (\hat{Z}_{\cdot j})^2 \\ &= \sum_{j=1}^J \left\{ \frac{1}{\sqrt{I}} \sum_{i=1}^I 2\sqrt{n_j} \left( \hat{\sigma}_{ij} - \sum_{k=1}^J \frac{n_k}{n} \hat{\sigma}_{ik} \right) \right\}^2 \\ &= \frac{4}{I} \sum_{j=1}^J n_j \left\{ \sum_{i=1}^I \left( \hat{\sigma}_{ij} - \sum_{k=1}^J \frac{n_k}{n} \hat{\sigma}_{ik} \right) \right\}^2 \end{aligned}$$

と書き換えられる.

$T_P$  の漸近分布は上式のようになる.

また統計量  $T_P$  の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  $\alpha$  の検定方式は

$T_P \geq \chi_{J-1}^2(\alpha)$  のとき  $H_0$  を棄却する

$T_P < \chi_{J-1}^2(\alpha)$  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  $\chi_{J-1}^2(\alpha)$  とおく

検関数  $\phi(x)$  を使って表すと,

$$\phi(x) = \begin{cases} 1 & (T_P \geq \chi_{J-1}^2(\alpha)) \\ 0 & (T_P < \chi_{J-1}^2(\alpha)) \end{cases}$$

である.

## 4 C 言語によるプログラム解説

### 4.1 プログラムの解説

C 言語により, カイ二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.(参考文献 [3] より). 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();
    keisan3();
    keisan4();
    keisan5();
}
```

```

XN=KAI(ALPHA,j);
printf("誤差 %f 以下の自由度 %d のカイ二乗分布の上側
%2.2f %点は %f \n", ERR, GU(j),100*ALPHA, XN);
printf("層別化された多群ポアソン分布の場合のカイ二
乗型検定\n");
if(T>XN){
    printf("H_0 を棄却する\n");
}
else{
    printf("H_0 を棄却しない\n");
}
return(0);
}

```

である.

## 4.2 プログラムの流れ

- 1.input 関数で, 標本数, 標本サイズ, 観測値を入力する.
- 2.GU 関数の中で自由度の値を出力する.
- 3.keisan1 関数で,  $\hat{\sigma}$  の値を計算する.
- 4.keisan2 関数で,  $\bar{c}$  の値を計算する.
- 5.keisan3 関数で, 統計量  $T_P$  の値を計算する.(段階 1)
- 6.keisan4 関数で, 統計量  $T_P$  の値を計算する.(段階 2)
- 7.keisan5 関数で, 統計量  $T_P$  の値を計算する.(段階 3)
- 8.main 関数で以上のプログラムを実行し, 有意水準  $\alpha$  の値を入力して, カイ二乗型検定の結果を出力する.

## 5 データとその解析結果

### 5.1 実行結果の例

本土 (北海道, 本州, 四国, 九州) において 1951 年から 2010 年の各年で, 7 月から 10 月までの上陸数と接近数のデータをもとに, 検定を行った. ここで, 台風の上陸とは, 中心が本土に達した場合のことをいい, 接近は中心がその地点を中心とする半径 300km 以内の域内に入ることをいう. (参考文献 [4] より).

このとき,  $I = 6$ (年代),  $J = 4$ (月),  $k = 300$ (各年代の日数) とする.

$\alpha=0.01$  のとき (1990 年代の上陸数, 接近数)

ディスプレイに表示される実行結果は次である.

T=11.202961

誤差 0.000010 以下の自由度 3 のカイ二乗分布の上側  
1.00 %点は 11.173836

層別化された多群ポアソン分布の場合のカイ二乗型検定  
H\_0 を棄却する

T=16.240944

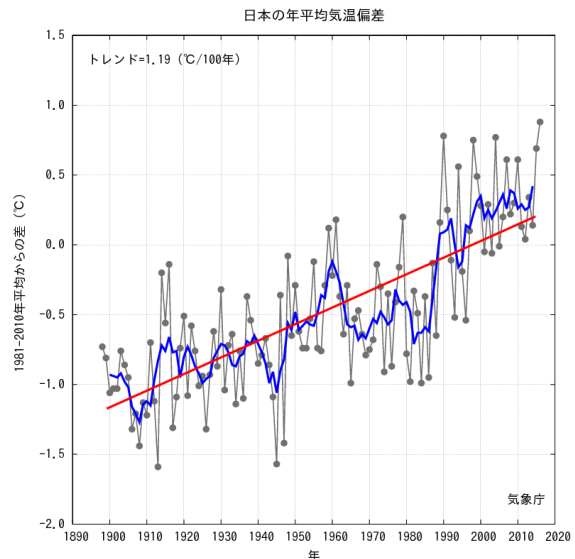
誤差 0.000010 以下の自由度 3 のカイ二乗分布の上側  
1.00 %点は 11.173836

層別化された多群ポアソン分布の場合のカイ二乗型検定  
H\_0 を棄却する

## 5.2 解析結果とその考察

検定を行った結果, 有意水準  $\alpha = 0.01$  について  $H_0$  を棄却した. この結果より, 1990 年代において 7,8,9,10 月における上陸数, 接近数は増加傾向にあることが分かった. 他にも棄却した年代がみられ, その理由として, 年平均気温との差が急激に広がっている年には上陸. 接近数は増加傾向にあると考察できる.

## 6 グラフ (参考文献 [5] より)



黒: 各年の平均気温の基準値からの偏差

青: 偏差の 5 年移動平均

赤: 長期的な変化傾向. 基準値は 1981~2010 年の 30 年平均

## 7 おわりに

本論では, 多群ポアソンモデルにおけるカイ二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同様の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ポアソン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] 白旗慎吾: 『統計解析入門』, 共立出版, 1992
- [2] 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
- [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』 南山大学大学院理工学研究科, 2015.
- [4] デジタル台風: 台風の上陸, 接近, 通過の定義, <http://agora.ex.nii.ac.jp/digitaltyphoon/help/landfall.html.ja>, 2018/1/18 閲覧
- [5] 気象庁: 日本の年平均気温, <http://www.data.jma.go.jp/cpdinfo/temp/anjpn.html>, 2018/1/18 閲覧