ガンマモデルにおけるダネット型多重比較法

2010SE121　宮崎　諒

指導教員：白石　高章

1　はじめに

本論では，ガンマ分布に従う多群モデルにおける Dunnett 型 (ダネット型) の多重比較検定と同時信頼区間を考察する．さらに，ダネット法によって解析を行う C 言語プログラムを作成し，鳥の寿命のデータの解析を行う．

2　ガンマ分布

確率変数 X の密度関数が，

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

となる分布をパラメータ (α, β) のガンマ分布といい，$Ga(\alpha, \beta)$ と表す．ただし，パラメータ α, β は正である．また，密度関数を微分して，平均 $E(X) = \alpha\beta$，分散 $V(X) = \alpha\beta^2$ とできる．そして，ガンマ分布は寿命の分布として，生物学・医学・薬学などの様々な分野に応用されている (白旗 [1])．

3　モデルの設定

白石 [2] の指数分布に従う多群モデルの設定を参考にして，ある要因 A があり，k 個の水準 $A_1, \cdots, A_k$ を考える．水準は群とも呼ばれる．水準 A_i における標本の観測値 $(X_{i1}, X_{i2}, \cdots, X_{in_i})$ は第 i 標本または第 i 群と呼ばれ，ガンマ分布 $Ga(\alpha_0, \beta_i)$ に従っているものとする．ただし，β_i は未知 ($\beta_1, \ldots, \beta_k$ はすべて未知) とする．すなわち，

$$f(x) = \frac{1}{\Gamma(\alpha_0)} x^{\alpha_0 - 1} e^{-\frac{x}{\beta_i}} (x > 0)$$

ただし，すべての X_{ij} は互いに独立であると仮定し，総標本サイズを $n \equiv n_1 + \cdots + n_k$(すべての観測値の個数) とおく．また，第 i 群の標本平均を $\bar{X}_i. \equiv \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$ とする．

表 1　k 群ガンマモデル

水準	群	データ	平均	分布
対照	第 1 群	$X_{11}, \cdots, X_{1n_1}$	$\alpha_0\beta_1$	$Ga(\alpha_0, \beta_1)$
処理 1	第 2 群	$X_{21}, \cdots, X_{2n_2}$	$\alpha_0\beta_2$	$Ga(\alpha_0, \beta_2)$
$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$	$\vdots$	$\vdots$
処理 $k-1$	第 k 群	$X_{k1}, \cdots, X_{kn_k}$	$\alpha_0\beta_k$	$Ga(\alpha_0, \beta_k)$

4　ダネット型の多重比較検定法

白石 [2] の指数分布に従う多群モデルの場合を参考にして，3 のモデルに対応して，第 1 群を対照群とするダネット型多重比較検定を論じる．第 1 群の対照群と第 i 群の処理群を比較することを考える．1 つの比較のための検定は，

$$\text{帰無仮説} \quad H_i: \quad \beta_i = \beta_1$$

に対して 3 種の対立仮説が

$$\text{(1) 両側対立仮説} \quad H_i^{A\pm}: \quad \beta_i \neq \beta_1$$

$$\text{(2) 片側対立仮説} \quad H_i^{A+}: \quad \beta_i > \beta_1$$

$$\text{(3) 片側対立仮説} \quad H_i^{A-}: \quad \beta_i < \beta_1$$

となる．ここで，

$$T_i \equiv \frac{\sqrt{\alpha_0}\{\log(\bar{X}_i.) - \log(\bar{X}_1.)\}}{\sqrt{\frac{1}{n_i} + \frac{1}{n_1}}}$$

とおく，また，漸近理論を述べるために，

(条件 1)

$$\lim_{n\to\infty} \frac{n_i}{n} = \lambda_i > 0 \ (1 \leq i \leq k)$$

を仮定する．このとき次の補題と定理を得る．

【補題 1】 $X_1, \cdots, X_n \sim Ga(\alpha_0, \beta)$ のとき，

$$\sqrt{n}\{\log(\bar{X}_n) - \log(\alpha_0\beta)\} \xrightarrow{L} N(0, \frac{1}{\alpha_0})$$

(証明) 中心極限定理とデルタ法 (白石 [3] の定理 3.27 と定理 3.35) を用いることで示すことができる．□

【定理 1】 (条件 1) の下で，

$$\lim_{n\to\infty} P_0\left(\max_{2\leq i\leq k} |T_i| \leq t\right) = B_1(t),$$

$$\lim_{n\to\infty} P_0\left(\max_{2\leq i\leq k} T_i \leq t\right) = B_2(t)$$

が成り立つ．ただし，

$$B_1(t) \equiv \int_{-\infty}^{\infty} \prod_{i=2}^{k} \left\{ \Phi\left(\sqrt{\frac{\lambda_i}{\lambda_1}} \cdot x + \sqrt{\frac{\lambda_i + \lambda_1}{\lambda_1}} \cdot t\right) - \Phi\left(\sqrt{\frac{\lambda_i}{\lambda_1}} \cdot x - \sqrt{\frac{\lambda_i + \lambda_1}{\lambda_1}} \cdot t\right) \right\} d\Phi(x),$$

$$B_2(t) \equiv \int_{-\infty}^{\infty} \prod_{i=2}^{k} \Phi\left(\sqrt{\frac{\lambda_i}{\lambda_1}} \cdot x + \sqrt{\frac{\lambda_i + \lambda_1}{\lambda_1}} \cdot t\right) d\Phi(x)$$

とする．

(証明) 補題 1 を変形すると，

$$\sqrt{\alpha_0}\sqrt{n}\{\log(\bar{X}_i.) - \log(\alpha_0\beta_i)\} \xrightarrow{L} Y_i \sim N\left(0, \frac{1}{\lambda_i}\right)$$

$$\sqrt{\alpha_0}\sqrt{n}\{\log(\bar{X}_1.)-\log(\alpha_0\beta_1)\}\xrightarrow{L}Y_1\sim N\left(0,\frac{1}{\lambda_1}\right)$$

ただし，$\bar{X}_i.\equiv\frac{1}{n_i}\sum_{j=1}^{n_i}X_{ij}$ とする．
ゆえに，(条件 1) を用いると，

$$\lim_{n\to\infty}P_0\left(\max_{2\le i\le k}|T_i|\le t\right)=P_0\left(\max_{2\le i\le k}\frac{|Y_i-Y_1|}{\sqrt{\frac{1}{\lambda_i}+\frac{1}{\lambda_1}}}\le t\right)$$

と変形できる．
次に白石 [4] の定理 A.7 より，示すことができる．□
ここで α を与え，
$B_1(t)=1-\alpha$ を満たす t の解を $b_1(k,\lambda_1,\cdots,\lambda_k;\alpha)$,
$B_2(t)=1-\alpha$ を満たす t の解を $b_2(k,\lambda_1,\cdots,\lambda_k;\alpha)$
とし，$b_1(k,\lambda_1,\cdots,\lambda_k;\alpha)$ を $b_1(\alpha)$, $b_2(k,\lambda_1,\cdots,\lambda_k;\alpha)$ を $b_2(\alpha)$ と表記する．このとき，(定理 1) より，次の漸近的な多重比較法が導かれる．
(対数変換を使った漸近的な多重比較法)
平均母数の制約に応じて，水準 α の漸近的な多重比較法は次のように与えられる．
両側検定 : 帰無仮説 $H_i vs.$対立仮説 ${H_i}^{A\pm}(i=2,\cdots,k)$ のとき，$|T_i|>b_1(\alpha)$ となる i に対して H_i を棄却し，対立仮説 ${H_i}^{A\pm}$ を受け入れ，$\beta_i\neq\beta_1$ と判定する．

ここで，$\boldsymbol{\beta}\equiv(\beta_1,\cdots,\beta_k)$ に対して，

$$T_i(\boldsymbol{\beta})\equiv\frac{\sqrt{\alpha_0}\{\log(\bar{X}_i.)-\log(\bar{X}_1.)\}}{\sqrt{\frac{1}{n_i}+\frac{1}{n_1}}}$$

とおく，このとき，次の定理を得る．
【定理 2】 (条件 1) の下で，

$$\lim_{n\to\infty}P_0\left(\max_{2\le i\le k}|T_i(\beta)|\le t\right)=B_1(t),$$

$$\lim_{n\to\infty}P_0\left(\max_{2\le i\le k}T_i(\beta)\le t\right)=B_2(t)$$

が成り立つ．ただし，

$$B_1(t)\equiv\int_{-\infty}^{\infty}\prod_{i=2}^{k}\Biggl\{\Phi\left(\sqrt{\frac{\lambda_i}{\lambda_1}}\cdot x+\sqrt{\frac{\lambda_i+\lambda_1}{\lambda_1}}\cdot t\right)$$
$$-\Phi\left(\sqrt{\frac{\lambda_i}{\lambda_1}}\cdot x-\sqrt{\frac{\lambda_i+\lambda_1}{\lambda_1}}\cdot t\right)\Biggr\}d\Phi(x),$$

$$B_2(t)\equiv\int_{-\infty}^{\infty}\prod_{i=2}^{k}\Phi\left(\sqrt{\frac{\lambda_i}{\lambda_1}}\cdot x+\sqrt{\frac{\lambda_i+\lambda_1}{\lambda_1}}\cdot t\right)d\Phi(x)$$

とする．
(証明) 定理 1 と同様に示すことができる．□
また，同様に，次の漸近的な同時信頼区間が導かれる．
(対数変換を使った漸近的な同時信頼区間)
$\log(\beta_i)-\log(\beta_1)$ に対する信頼係数 $1-\alpha$ の同時信頼区間は，次のように与えられる．ただし，$i=2,\cdots,k$ とする．
両側信頼区間

$$\log(\bar{X}_i.)-\log(\bar{X}_1.)-b_1(\alpha)\cdot\sqrt{\frac{1}{n_i}+\frac{1}{n_1}}$$
$$<\log(\beta_i)-\log(\beta_1)$$
$$<\log(\bar{X}_i.)-\log(\bar{X}_1.)+b_1(\alpha)\cdot\sqrt{\frac{1}{n_i}+\frac{1}{n_1}}$$

5 プログラムの流れ

大橋 [5] を参考に，ダネット法によって解析を行う C 言語プログラムを作成した．プログラムの手順は次のようになる．
1. teigi1()，teigi2() において，$b_1(k,\lambda_1,\cdots,\lambda_k;\alpha)$, $b_2(k,\lambda_1,\cdots,\lambda_k;\alpha)$ の値の配列を用意する．
2. input() において，群の個数 k，ガンマ分布のパラメータ α_0，有意水準 α，検定の範囲を入力する．
3. hantei() において，2 の値で検定が行えるかどうかを確認する．
4. output() において，$b_1(k,\lambda_1,\cdots,\lambda_k;\alpha)$, $b_2(k,\lambda_1,\cdots,\lambda_k;\alpha)$ の値を選択し，出力する．
5. input2() において，各群のサイズとそれぞれの観測値を入力する．
6. keisan() において，T_i を計算する．
7. kekka() において，検定結果と信頼区間を出力する．

6 おわりに

本論では，ガンマ分布に従う多群モデルにおける多重比較検定と同時信頼区間を考察し，C 言語プログラムを作成して鳥の寿命の解析を行った．その結果から，鳥の寿命は全長の大きい鳥ほど長いということが分かった．また，ガンマモデルにおけるダネット型多重比較法の有用性を改めて確認することができた．

参考文献

[1] 白旗慎吾．「統計学」．ミネルヴァ書房．2008

[2] 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．34.1-20

[3] 白石高章．「統計学の基礎-データと結びつきがよくわかる数理」．日本評論社．2012

[4] 白石高章．「多群連続モデルにおける多重比較法-パラメトリック, ノンパラメトリックの数理統計」．共立出版．2011

[5] 大橋一之 : すべての平均相違の多重比較統計量の分布論．2012 年度南山大学情報理工学部情報システム数理学科卒業論文，2013.