多群ワイブルモデルにおける順序制約のある場合の多重比較法

M2014SS004　宮崎諒

指導教員：白石高章

1　はじめに

本論では，ワイブル分布に従う多群モデルにおける順序制約のある場合の多重比較検定を考察する．多重比較法には様々なものがある．例えば，Hayter 法 (ヘイター法)，Tukey-Kramer 法 (テューキー・クレーマー法)，Dunnett 法 (ダネット法) などがある．本論ではヘイター法に類似した手法と閉検定手順による多重比較検定について考察する．また，それぞれの多重比較法の C 言語プログラムを作成し，データ解析を行う．

2　ワイブル分布

確率変数 X の密度関数が

$$f(x)=\frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1}\exp\left\{-\left(\frac{x}{\eta}\right)^{\beta}\right\}\qquad (x>0)$$

となる分布をパラメータ (β,η) のワイブル分布といい，$We(\beta,\eta)$ と表す．ただし，パラメータ β,η は正である．$\beta=1$ のときにワイブル分布は指数分布となる．この β を shape parameter (形状パラメータ)，η を scale parameter (尺度パラメータ) とも呼ぶ．また，ワイブル分布の平均と分散はそれぞれ

$$E(X)=\eta\Gamma\left(1+\frac{1}{\beta}\right)$$

$$V(X)=\eta^2\left\{\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^2\left(1+\frac{1}{\beta}\right)\right\}$$

とできる．そして，ワイブル分布は医学・工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] の正規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，ワイブル分布 $We(\beta_0,\eta_i)$ に従っているものとする．ただし，$1\le i\le k$ とし，η_i は未知 ($\eta_1,\ldots,\eta_k$ はすべて未知) とする．すなわち，密度関数は

$$f(x)=\frac{\beta_0}{\eta_i}\left(\frac{x}{\eta_i}\right)^{\beta_0-1}\exp\left\{-\left(\frac{x}{\eta_i}\right)^{\beta_0}\right\}\qquad (x>0)$$

となる．また，本論では

$$\gamma(i)\equiv\Gamma\left(1+\frac{i}{\beta_0}\right)\qquad (i=1,2)$$

と定義し，平均と分散はこれを用いて，$E(X)=\eta_i\gamma(1)$，$V(X)={\eta_i}^2\{\gamma(2)-\gamma^2(1)\}$ とできる．また，本論では $\eta_1,\cdots,\eta_k$ に傾向性の制約

(制約 1)

$$\eta_1\le\eta_2\le\cdots\le\eta_k$$

がある場合を考える．ゆえに，標本サイズが等しいという制約

(制約 2)

$$n_1=\cdots=n_k=n_0$$

を付加する．よって，総標本サイズ n (すべての観測値の個数) は，$n\equiv n_1+\cdots+n_k=kn_0$ とできる．ただし，すべての X_{ij} は互いに独立であると仮定し，第 i 群の標本平均を $\bar{X}_{i\cdot}\equiv\frac{1}{n_0}\sum_{j=1}^{n_0}X_{ij}$ とする．このとき，次の k 群ワイブルモデルの表 1 を得る．さらに，本論では各群のパラメータの値 β_i,η_i を次の 2 つの式を用いて推定し，β_0 については β_i の平均を求めることで値が得られる (蓑谷 [3])．

$$\beta_i=\left\{\left(\sum_{j=1}^{n_0}X_j\log X_j\right)\left(\sum_{j=1}^{n_0}X_j\right)^{-1}-\frac{1}{n_0}\sum_{j=1}^{n_0}\log X_j\right\}^{-1}$$

$$\eta_i=\left(\frac{1}{n_0}\sum_{j=1}^{n_0}{X_j}^{\beta_i}\right)^{1/\beta_i}$$

表 1　k 群ワイブルモデル

群	データ	平均	分布
第 1 群	$X_{11},\cdots,X_{1n_1}$	$\eta_1\gamma(1)$	$We(\beta_0,\eta_1)$
第 2 群	$X_{21},\cdots,X_{2n_2}$	$\eta_2\gamma(2)$	$We(\beta_0,\eta_2)$
$\vdots$	$\vdots\quad\vdots\quad\vdots$	$\vdots$	$\vdots$
第 k 群	$X_{k1},\cdots,X_{kn_k}$	$\eta_k\gamma(k)$	$We(\beta_0,\eta_k)$

4　分散安定化変換による漸近理論

(制約 1)，(制約 2) の下で，次の補題と定理を得る．

【補題 1】　(制約 2) の下で次の式が成り立つ．

$$\sqrt{\frac{\gamma^2(1)}{\gamma(2)-\gamma^2(1)}}\sqrt{n}\{\log(\bar{X}_{i\cdot})-\log(\eta_i\gamma(1))\}\xrightarrow{L}N(0,k)\quad(1)$$

(証明)　(制約 2) の下で鬼頭 [4] の補題 1，補題 2 より (1) 式が得られる．□

【定理 1】　$1\le i<i'\le k$ とし，

$$T_{i'i}(\boldsymbol{\eta})\equiv\sqrt{\frac{\gamma^2(1)}{\gamma(2)-\gamma^2(1)}}\sqrt{\frac{n_0}{2}}\big[\log(\bar{X}_{i'\cdot})-\log(\bar{X}_{i\cdot})-\{\log(\eta_{i'}\gamma(1))-\log(\eta_i\gamma(1))\}\big]\,,$$

$$T_{i'i} \equiv \sqrt{\frac{\gamma^2(1)}{\gamma(2)-\gamma^2(1)}}\sqrt{\frac{n_0}{2}}\{\log(\bar{X}_{i'}.)-\log(\bar{X}_{i}.)\}$$

とすると，(制約 1)，(制約 2) の下で

$$\lim_{n\to\infty} P\left(\max_{1\le i<i'\le k} T_{i'i}(\boldsymbol{\eta}) \le t\right)$$
$$= \lim_{n\to\infty} P_0\left(\max_{1\le i<i'\le k} T_{i'i} \le t\right) = C_1(t) \qquad (2)$$

が成り立つ．ただし，

$$C_1(t) \equiv P\left(\max_{1\le i<i'\le k} \frac{Z_{i'}-Z_i}{\sqrt{2}} \le t\right)$$

とし，Z_i は独立で同一の標準正規分布 $N(0,1)$ に従う確率変数とする．

(証明)　補題 1 より，

$$\lim_{n\to\infty} P\left(\max_{1\le i<i'\le k} T_{i'i}(\boldsymbol{\eta}) \le t\right)$$
$$= \lim_{n\to\infty} P\Bigg(\max_{1\le i<i'\le k} \sqrt{\frac{\gamma^2(1)}{\gamma(2)-\gamma^2(1)}}\sqrt{\frac{n_0}{2}}$$
$$\big[\log(\bar{X}_{i'}.)-\log(\bar{X}_{i}.)-\{\log(\beta_{i'}\gamma(1))-\log(\beta_i\gamma(1))\}\big] \le t\Bigg)$$
$$= \lim_{n\to\infty} P\Bigg(\max_{1\le i<i'\le k} \frac{1}{\sqrt{2}}\sqrt{\frac{\gamma^2(1)}{\gamma(2)-\gamma^2(1)}}\sqrt{\frac{n_0}{n}}\sqrt{n}$$
$$\big[\log(\bar{X}_{i'}.)-\log(\bar{X}_{i}.)-\{\log(\eta_{i'}\gamma(1))-\log(\eta_i\gamma(1))\}\big] \le t\Bigg)$$
$$= P\left(\max_{1\le i<i'\le k} \frac{Z_{i'}-Z_i}{\sqrt{2}} \le t\right)$$
$$= C_1(t)$$

となり，同様に，

$$\lim_{n\to\infty} P_0\left(\max_{1\le i<i'\le k} T_{i'i} \le t\right) = C_1(t)$$

も示すことができるので，(2) 式が得られる．□

5　多重比較検定

白石 [2] の正規分布に従う多群モデルの場合を参考にして，3 節のモデルに対応する多重比較検定を論じる．(制約 1) の下で，k 個の水準のすべての比較を考える．一つの比較のための検定は

帰無仮説 $H_{(i,i')}\colon \eta_i = \eta_{i'}$ $vs.$ 対立仮説 $H^A_{(i,i')}\colon \eta_i < \eta_{i'}$

となり，帰無仮説のファミリーを

$$\mathcal{H}_1 \equiv \{H_{(1,2)},\cdots,H_{(1,k)},H_{(2,3)},\cdots,H_{(2,k)},\cdots,H_{(k-1,k)}\}$$
$$= \{H_{(i,i')} \mid 1\le i<i'\le k\}$$

とおく．

5.1　ヘイター法に類似した手法による多重比較検定

ヘイター法に類似した手法による多重比較検定を論じるために，Hayter and Liu [5] を参考にして，次の補題を示す．

【補題 2】　$C_1(t)$ は次の漸化式によって与えられる．

$$h_1(t,y) = \Phi\left(\sqrt{2}\cdot t + y\right) \qquad (3)$$

$$h_r(t,y) = \int_{-\infty}^{y} h_{r-1}(t,x)\varphi(x)dx$$
$$+ h_{r-1}(t,y)\left\{\Phi\left(\sqrt{2}\cdot t+y\right)-\Phi\Big(y\Big)\right\} \quad (r\ge 2) \quad (4)$$

$$C_1(t) = \int_{-\infty}^{\infty} h_{k-1}(t,y)\varphi(y)dy \qquad (5)$$

ただし，$\varphi(x)$ と $\Phi(x)$ をそれぞれ標準正規分布の密度関数，分布関数とする．

(証明)　$h_1(t,y)$ と $h_r(t,y)$ $(r\ge 2)$ を次のように定義すると，

$$h_1(t,y) = P\{Z_1 - y \le \sqrt{2}\cdot t\} = \Phi\left(\sqrt{2}\cdot t + y\right)$$

$$h_r(t,y) = P\left\{\max_{1\le i<i'\le r+1}(Z_i - Z_{i'}) \le \sqrt{2}\cdot t \;\Big|\; Z_{r+1}=y\right\}$$
$$= P\left\{\max_{1\le i<i'\le r}(Z_i - Z_{i'}) \le \sqrt{2}\cdot t,\ \max_{1\le i\le r}(Z_i - y) \le \sqrt{2}\cdot t\right\}$$
$$= \int_{-\infty}^{\sqrt{2}\cdot t+y} \varphi(x)P\Bigg\{\max_{1\le i<i'\le r}(Z_i - Z_{i'}) \le \sqrt{2}\cdot t,$$
$$\max_{1\le i\le r-1}(Z_i - y) \le \sqrt{2}\cdot t \;\Big|\; Z_r = x\Bigg\}dx$$
$$= \int_{-\infty}^{\sqrt{2}\cdot t+y} \varphi(x)P\Bigg\{\max_{1\le i<i'\le r-1}(Z_i - Z_{i'}) \le \sqrt{2}\cdot t,$$
$$\max_{1\le i\le r-1} Z_i \le \min\left\{\sqrt{2}\cdot t+y, \sqrt{2}\cdot t+x\right\}\Bigg\}dx$$
$$= \int_{-\infty}^{y} h_{r-1}(t,x)\varphi(x)dx$$
$$+ h_{r-1}(t,y)\left\{\Phi\left(\sqrt{2}\cdot t+y\right)-\Phi\Big(y\Big)\right\}$$

とでき，(3) 式と (4) 式が導かれる．(4) 式を再帰的に計算をすると (5) 式を得ることができる．□

ここで，白石 [2] の正規分布に従う多群モデルの場合を参考にして，3 節のモデルに対応するヘイター法に類似した手法よる多重比較検定を論じる．補題 2 より，α $(0<\alpha<1)$ を与え，$C_1(t) = 1-\alpha$ を満たす t の解を $c_1(k;\alpha)$ とする．このとき，定理 1 より，次の漸近的な多重比較検定と同時信頼区間が導かれる．

[対数変換を使った漸近的な多重比較検定]
{ 帰無仮説 $H_{(i,i')}$ vs. 対立仮説 $H^A_{(i,i')}|1 \le i < i' \le k\}$ に対する水準 α の漸近的な多重比較検定は次のように与えられる．
(1) $i < i'$ となるペア i, i' に対して $T_{i'i} > c_1(k;\alpha)$ ならば，帰無仮説 $H_{(i,i')}$ を棄却し，対立仮説 $H^A_{(i,i')}$ を受け入れ，$\eta_i < \eta_{i'}$ と判定する．
(2) $i < i'$ となるペア i, i' に対して $T_{i'i} < c_1(k;\alpha)$ ならば，帰無仮説 $H_{(i,i')}$ を棄却しない．

[対数変換を使った漸近的な同時信頼区間]
$\log(\eta_{i'}) - \log(\eta_i)$ に対する信頼係数 $1-\alpha$ の漸近的な同時信頼区間は，次のように与えられる．ただし，$(1 \le i < i' \le k)$ とする．

$$\log(\bar{X}_{i'\cdot}) - \log(\bar{X}_{i\cdot}) - c_1(k;\alpha)\sqrt{\frac{2}{n_0}}\sqrt{\frac{\gamma(2)-\gamma^2(1)}{\gamma^2(1)}}$$
$$< \log(\eta_{i'}) - \log(\eta_i) < \infty$$

5.2 閉検定手順による多重比較検定

白石 [2] の正規分布に従う多群モデルの場合を参考にして，3 節のモデルに対応する閉検定手順による多重比較検定を論じる．$\mathcal{U}_1 \equiv \{(i,i') \mid 1 \le i < i' \le k\}$ とおき，$\mathcal{H}_1$ の要素の仮説 $H_{(i,i')}$ の論理積からなるすべての集合は

$$\overline{\mathcal{H}}_1 \equiv \left\{ \bigwedge_{\boldsymbol{v}\in V} H_{\boldsymbol{v}} \;\middle|\; \emptyset \subset V \subset \mathcal{U}_1 \ (V \ne \emptyset) \right\}$$

で表され，$\overline{\mathcal{H}}_1$ は $\mathcal{H}_1$ の閉包と呼ばれる．また，$\emptyset \subset V \subset \mathcal{U}_1\ (V \ne \emptyset)$ を満たす任意の V に対して，

$$\bigwedge_{\boldsymbol{v}\in V} H_{\boldsymbol{v}} : 任意の\ (i,i') \in V\ に対して, \eta_i = \eta_{i'}$$

という仮説が得られる．$I_1, \cdots, I_J\ (I_j \ne \emptyset, j = 1, \cdots, J)$ を，次の性質を満たす添え字 $\{1, \cdots, k\}$ の互いに素な部分の集合の組とする．

(性質 1)
ある整数 $\ell_1, \cdots, \ell_J \ge 2$ とある整数 $0 \le s_1 < \cdots < s_J < k$ が存在して，$I_j = \{s_j+1, s_j+2, \cdots, s_j+\ell_j\}$ $(j = 1, \cdots, J)$, $s_j+\ell_j \le s_{j+1}$ $(j = 1, \cdots, J-1)$ かつ $s_J+\ell_J \le k$ が成り立つ．ただし，I_j は連続した整数の要素からなり，$\ell_j = \#(I_j) \ge 2$ であり，$\#(I_j)$ は集合 I_j の要素の個数を表す．

ここで，同じ $I_j\ (j = 1, \cdots, J)$ に含まれる添え字をもつという場合における帰無仮説を $H(I_1, \cdots, I_J)$ と表す．このとき，$\emptyset \subset V \subset \mathcal{U}_1\ (V \ne \emptyset)$ を満たす任意の V に対して，(性質 1) で述べたある自然数 J と上記の $I_1, \cdots, I_J$ が存在して，

$$\bigwedge_{\boldsymbol{v}\in V} H_{\boldsymbol{v}} = H(I_1, \cdots, I_J)$$

が成り立つ．さらに，仮説 $H(I_1, \cdots, I_J)$ は

$$H(I_1, \cdots, I_J) : \eta_{s_j+1} = \eta_{s_j+2} = \cdots = \eta_{s_j+\ell_j}$$

と表現できる．また，定理 1 の $T_{i'i}$ を用いて，

$$T(I_j) \equiv \max_{s_j+1 \le i < i' \le s_j+\ell_j} T_{i'i}$$

とおき，$H(I_1, \cdots, I_J)$ に対して，M を

$$M \equiv M(I_1, \cdots, I_J) \equiv \sum_{j=1}^{J} \ell_j$$

とする．さらに，$\ell \le k$ となる自然数 ℓ に対して，

$$C_1(t|\ell) \equiv P\left(\max_{1 \le i < i' \le \ell} \frac{Z_{i'} - Z_i}{\sqrt{2}} \le t \right)$$

とおく．そこで $\alpha\ (0 < \alpha < 1)$ を与え，$C_1(t|\ell) = 1-\alpha$ を満たす t の解を $c_1(\ell;\alpha)$ とする．このとき，次の閉検定手順が導かれる．

[閉検定手順]
(a) $J \ge 2$ のとき，$\ell = \ell_1, \cdots, \ell_J$ に対して，

$$\alpha(M,\ell) \equiv 1 - (1-\alpha)^{\ell/M}$$

で $\alpha(M,\ell)$ を定義する．$1 \le j \le J$ となるある整数 j が存在して，

$$c_1(\ell_j;\alpha(M,\ell_j)) < T(I_j)\ ならば帰無仮説\ \bigwedge_{\boldsymbol{v}\in V} H_{\boldsymbol{v}} を棄却する．$$

(b) $J = 1\ (M = \ell_1)$ のとき，

$$c_1(M;\alpha) < T(I_1)\ ならば帰無仮説\ \bigwedge_{\boldsymbol{v}\in V} H_{\boldsymbol{v}} を棄却する．$$

(a)，(b) の方法で，$(i,i') \in V \subset \mathcal{U}_1$ を満たす任意の V に対して，$\bigwedge_{\boldsymbol{v}\in V} H_{\boldsymbol{v}}$ が棄却されるとき，多重比較検定として，$H_{(i,i')}$ を棄却する．

ここで，(a)，(b) の検定の有意水準は漸近的に α であることを次の定理 2 に示す．
【定理 2】 提案した閉検定手順は漸近的に水準 α の多重比較検定である．
(証明) (b) の検定の有意水準が α であることは自明であるので，(a) の検定の有意水準も α であることを示す．$T(I_1), \cdots, T(I_J)$ は互いに独立なので，

$$\lim_{n\to\infty} P_0(T(I_j) \le c_1(\ell_j;\alpha(M,\ell_j)), j = 1, \cdots, J)$$
$$= \prod_{j=1}^{J} \{\lim_{n\to\infty} P_0(T(I_j) \le c_1(\ell_j;\alpha(M,\ell_j))\}$$

が成り立つ．ここで，白石 [6] の定理 A.5 より，

$$(与式) \ge \prod_{j=1}^{J} C_1(c_1(\ell_j;\alpha(M,\ell_j))|\ell_j)$$

$$= \prod_{j=1}^{J}(1-\alpha(M,\ell_j))$$

$$= \prod_{j=1}^{J}\{(1-\alpha)^{\ell_j/M}\}$$

$$= 1-\alpha$$

を得る．よって，

$$\lim_{n\to\infty} P_0(\text{ある } j \text{ が存在して，} T(I_j) > c_1(\ell_j;\alpha(M,\ell_j)))$$

$$= 1-\lim_{n\to\infty} P_0(T(I_j) \le c_1(\ell_j;\alpha(M,\ell_j)), j=1,\cdots,J)$$

$$\le \alpha$$

が成り立つ．したがって，(a) の検定の有意水準は漸近的に α である．□

ここで，$k=5$ のとき，$H(I_1,\cdots,I_J) \in \overline{\mathcal{H}}_{(1,2)}$ とした場合を例として考える．閉検定手順により多重比較検定として，帰無仮説 $H_{(1,2)}$ が棄却される場合に，検定される帰無仮説 $H(I_1,\cdots,I_J)$ を表 2 として挙げる．この表から，$H_{(1,2)}$ が多重比較検定として棄却されるためには 8 個の帰無仮説を棄却しなければならない．

表 2　$k=5$ のときの $H(I_1,\cdots,I_J) \in \overline{\mathcal{H}}_{(1,2)}$

M	$H(I_1,\cdots,I_J)$
5	$H(\{1,2,3,4,5\}), H(\{1,2,3\},\{4,5\}), H(\{1,2\},\{3,4,5\})$
4	$H(\{1,2,3,4\}), H(\{1,2\},\{3,4\}), H(\{1,2\},\{4,5\})$
3	$H(\{1,2,3\})$
2	$H(\{1,2\})$

6　データ解析

ヘイター法に類似した手法と閉検定手順による多重比較検定を行う C 言語プログラムを作成し，肺がん患者の生存時間とカルノフスキー評点のデータ (蓑谷 [7]) を用いて，データ解析を行う．プログラムの流れは以下の通りである。

(ヘイター法に類似した手法のプログラムの流れ)
1. teigi() において，付表 B.1，付表 B.2 の $c_1(k;\alpha)$ の値を配列として用意する．
2. input() において，群の個数 k，ワイブル分布の shape parameter β_0，有意水準 α を入力する．
3. hantei() において，手順 2 で入力した値で多重比較検定が行えるかどうかを確認する．
4. output() において，手順 1 で用意した $c_1(k;\alpha)$ の値を選択し，出力する．
5. input2() において，群のサイズとデータの観測値を入力する．
6. keisan() において，$T_{i'i}$ を計算する．
7. kekka() において，多重比較検定の結果と同時信頼区間を出力する．

(閉検定手順のプログラムの流れ)
1. teigi() において，付表 B.1，付表 B.2 の $c_1(\ell;\alpha(M,\ell))$ 及び $c_1(k;\alpha)$ の値を配列として用意する．
2. input() において，群の個数 k，ワイブル分布の shape parameter β_0，有意水準 α，検定する群 i,i' を入力する．
3. hantei() において，手順 2 で入力した値で多重比較検定が行えるかどうかを確認する．
4. input2() において，群のサイズとデータの観測値を入力する．
5. keisan() において，$T_{i'i}$ を計算する．
6. kentei() において，すべての閉検定手順を出力する．
7. kekka13()〜kekka55() において，多重比較検定の結果と同時信頼区間を出力する．

また，データ解析を行うために，ワイブル分布のパラメータを推定するための C 言語プログラムも作成する．データ解析の結果から，肺がん患者の生存時間とカルノフスキー評点には深い関係性があり，カルノフスキー評点の有用性を確認できる．また，ヘイター法に類似した手法よりも閉検定手順の方が検出力が高いことも確認できる．

7　おわりに

本論では，ワイブル分布に従う多群モデルにおける順序制約のある場合の多重比較法を考察した．さらに，C 言語プログラムによるデータ解析を行った．C 言語プログラムによるデータ解析を通して，多重比較法の研究だけでなく，C 言語に関する研究も行うことができた．

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．多群連続モデルにおける位置母数に順序制約のある場合の閉検定手順．日本統計学会誌．43，215-245．2014

[3] 蓑谷千凰彦:『統計分布ハンドブック』．朝倉書店．2003

[4] 鬼頭広大．多群ワイブルモデルにおけるすべての尺度母数相違の多重比較法．2015 年度南山大学大学院理工学研究科システム数理専攻修士論文要旨．2016.1

[5] Hayter, A. J. and Liu, W．Exact calculations for the one-sided studentized range test for testing against a simple ordered alternative．Computational Statistics & Data Analysis．22，17-25．1996

[6] 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

[7] 蓑谷千凰彦 :『一般化線形モデルと生存分析』．朝倉書店．2013