すべての平均相違の多重比較統計量の分布論

2009SE221 大橋一之

指導教員：白石高章

1　はじめに

多重比較法とは, 3 つ以上の群において, 主に 2 つの群の間に違いがあるかを検出する手法である. 本研究の目的は, すべての平均相違の多重比較統計量の分布を簡素な確率変数で表現することにある. そのためにいくつかの命題を自分で導いた.

2　モデルの定義

次のような各群が同一の連続型分布関数 $F(x-\mu_i)$ をもつ次の k 群モデルを設定する. 尚, 本研究に用いるモデルは参考文献 [1] を参照にして記述した.

群	サイズ	データ	平均	分布関数
第 1 群	n_1	$X_{11},\cdots,X_{1n_1}$	μ_1	$F(x-\mu_1)$
第 2 群	n_2	$X_{21},\cdots,X_{2n_2}$	μ_2	$F(x-\mu_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
第 k 群	n_k	$X_{k1},\cdots,X_{kn_k}$	μ_k	$F(x-\mu_k)$

$$\mu_i=E(X_i),\ \ P(X_{ij}\le x)=F(x-\mu_i),$$
$$0<\sigma^2\equiv V(X_{ij})<\infty$$

これらの群の総標本サイズは $n\equiv n_1+\cdots+n_k$ とする. 続いて, 以下のような分布関数を定義する.

$$TA(t)\equiv k\int_0^\infty\left[\int_{-\infty}^\infty\left\{\Phi(x)-\Phi(x-\sqrt{2}\cdot ts)\right\}^{k-1}d\Phi(x)\right]g(s)ds,$$

$$TA^*(t)\equiv\sum_{j=1}^k\int_0^\infty\left[\int_{-\infty}^\infty\prod_{\substack{i=1\\i\neq j}}\left\{\Phi\left(\sqrt{\frac{\lambda_{n_i}}{\lambda_{n_j}}}\cdot x\right)-\Phi\left(\sqrt{\frac{\lambda_{ni}}{\lambda_{nj}}}\cdot x-\sqrt{\frac{\lambda_{ni}+\lambda_{nj}}{\lambda_{nj}}}\cdot ts\right)\right\}d\Phi(x)\right]g(s)ds$$

ただし,

$$\lambda_{ni}\equiv\frac{n_i}{n}(i=1,\cdots,k),$$
$$g(s)\equiv\frac{m^{m/2}}{\Gamma(m/2)2^{(m/2-1)}}s^{m-1}\exp(-ms^2/2)=\frac{mse^{-s^2}c^{(m/2-1)}}{\Gamma(m/2)},$$
$$c\equiv ms^2e^{-s^2}/2,\ \ m\equiv n-k$$

とする. また, $TA(t)$, $TA^*(t)$ に関連した分布として

$$A(t)\equiv k\int_{-\infty}^\infty\left\{\Phi(x)-\Phi(x-\sqrt{2}\cdot t)\right\}^{k-1}d\Phi(x)$$

とする.

3　多重比較法

文献 [2], [3] のテューキー・クレーマー法はすべての平均相違である多重比較法である. その統計量は,

$$T_{ij}\equiv\frac{\bar{X}_{i\cdot}-\bar{X}_{j\cdot}-\mu_i+\mu_j}{\sqrt{V_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}}\quad(1\le i<j\le k)$$

である. ただし, $V_E\equiv\dfrac{1}{n-k}\displaystyle\sum_{i=1}^k\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i\cdot})^2$ とする. このとき, 次の命題 1,2 を得る.

命題 1　$F(x)$ は $N(0,\sigma^2)$ の分布関数とする. $U\equiv mV_E/\sigma^2\sim\chi^2_m$ としたとき, 任意の t に対して,

$$P\left(\max_{1\le i<j\le k}|T_{ij}|\le t\right)=P\left(\max_{1\le i<j\le k}\frac{|Y_i-Y_j|}{\sqrt{\frac{U}{m}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}}\le t\right),$$

$$TA^*(t)=\sum_{j=1}^kP\left(Y_j-t\sqrt{\frac{U}{m}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}\le Y_i\le Y_j,\ i=1,2,\cdots,k,i\neq j\right)$$

が成り立つ. また, $n_1=\cdots=n_k=n_0$ のとき,

$$TA(t)=P\left(\max_{1\le i<j\le k}\frac{|Z_i-Z_j|}{\sqrt{\frac{2U}{m}}}\le t\right)$$

が成り立つ. ただし,

$$Y_i\equiv\frac{\bar{X}_{i\cdot}}{\sigma}\sim N\left(0,\frac{1}{n_i}\right),\ \ Z_i\equiv\frac{\sqrt{n_0}\bar{X}_{i\cdot}}{\sigma}\sim N(0,1)$$

であり, $Z_1,\cdots,Z_k$, U は互いに独立である. □

命題 2　$F(x)$ は $N(0,\sigma^2)$ の分布関数とする. このとき,

$$\lim_{n\to\infty}TA(t)=A(t)$$

が成り立つ. また,

$$\lim_{m\to\infty}TA^*(t)=\sum_{j=1}^kP\left(Y_j-t\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}\le Y_i\le Y_j,\ i=1,2,\cdots,k,i\neq j\right)$$

$$\lim_{m\to\infty}P\left(\max_{1\le i<j\le k}|T_{ij}|\le t\right)=P\left(\max_{1\le i<j\le k}\frac{|Y_i-Y_j|}{\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}}\le t\right)\tag{1}$$

である. □

$F(x)$ が正規分布の分布関数でない場合の分布論を述べる.

補題 1 $F(x)$ は一般の連続型分布関数とする.

$$\lim_{n\to\infty} n_i/n = \lambda_i > 0 \qquad (i = 1, \cdots, k)$$

としたとき,

$$T_{ij} \xrightarrow{\mathcal{L}} N(0,1)$$

が成り立つ. また,

$$\max_{1\le i<j\le k} |T_{ij}| \xrightarrow{\mathcal{L}} \max_{1\le i<j\le k} \frac{|Y_i - Y_j|}{\sqrt{\frac{1}{\lambda_i}+\frac{1}{\lambda_j}}}$$

である. ただし, $Y_i \sim N(0, 1/\lambda_i)$ である.

補題 1 を使って, 次の命題 3 を得る.

命題３ $F(x)$ は一般の連続型分布関数とし, $\lim_{n\to\infty} n_i/n = \lambda_i > 0$ とする.

任意の $t > 0$ に対して, $G_n(t) = P\left(\max_{1\le i<j\le k} |T_{ij}| \le t\right)$ とするとき,

$$\lim_{n\to\infty} G_n(t) = P\left(\max_{1\le i<j\le k} \frac{|Y_i - Y_j|}{\sqrt{\frac{1}{\lambda_i}+\frac{1}{\lambda_j}}} \le t\right),$$

$$A(t) \le P\left(\max_{1\le i<j\le k} \frac{|Y_i - Y_j|}{\sqrt{\frac{1}{\lambda_i}+\frac{1}{\lambda_j}}} \le t\right) \le A^*(t)$$

の関係が成り立つ. ただし,

$$A^*(t) \equiv \sum_{j=1}^{k} \int_{-\infty}^{\infty} \prod_{\substack{i=1\\ i\ne j}} \Bigg\{ \Phi\left(\sqrt{\frac{\lambda_i}{\lambda_j}} \cdot x\right) - \Phi\left(\sqrt{\frac{\lambda_i}{\lambda_j}} \cdot x - \sqrt{\frac{\lambda_i+\lambda_j}{\lambda_j}} \cdot t\right)\Bigg\} d\Phi(x),$$

$Y_i \sim N(0, 1/\lambda_i)$ である. □

命題 3 は文献 [4] の順位統計量の漸近結果と同じである.

4 テューキー・クレーマーの方法の実行プログラム

前述で述べたテューキー・クレーマーの方法を行うプログラムをＣ言語で作成した. パラメータの制約は $3 \le k \le 10$, $5 \le m \le 200$, $\alpha = 0.01, 0.05$ とする. $200 < m$ の場合はすべての平均相違の多重比較統計量の漸近分布の上側 $100\alpha\%$ 点を使用するものとする.

4.1 主プログラム

ページ数の都合上, 主プログラムのみをプログラムの実行手順に沿って示す. 全体のプログラムは卒業論文を参照せよ.

1. teigi(); にて, $ta(k, m; \alpha)$, $a(k; \alpha)$ の値を予め用意した配列に格納する.
2. input(); にて群の個数 k, 自由度 m, 有意水準 α の値を入力する.
3. hantei(); にて k, m, α の値がプログラムにおける許容範囲か確認する. 許容範囲外であるとき, メッセージを表示して手順 2 に戻る.
4. output(); にて, 値に沿った $ta(k, m; \alpha)$ の値を出力する. 尚, $m > 200$ の場合, $a(k; \alpha)$ の値を出力する.
5. input2(); にて各群のサイズとそれぞれの観測値を入力する.
6. keisan(); にて, 検定統計量 T_{ij} を算出する.
7. kekka(); にて検定結果と信頼区間を出力する.

ただし, $ta(k, m; \alpha)$ は方程式 $TA(t) = 1 - \alpha$ を満たす t の解であり, $a(k; \alpha)$ は $A(t) = 1 - \alpha$ を満たす t の解である. 尚, $ta(k, m; \alpha)$, の値は参考文献 [5] の表 6.1,6.2 をプログラムファイルに転写し, $a(k; \alpha)$ の値は参考文献 [1] の付表 12 を転写した.

5 おわりに

本稿では, テューキー・クレーマー法に関して, 統計量分布を簡素な確率変数で表現する方法を探った. また, テューキークレーマー法によるデータ解析や実行プログラムの作成を行うことで, これまでに研究した分野の実用性を感じることが出来た.

参考文献

[1] 白石高章 (2011).『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』. 共立出版株式会社，東京.

[2] Kramer, C.Y.(1956). Extension of multiple range tests to goup means with unequal numbers of replications. *Biometrics*, **8.** 75-86.

[3] Tukey, J.W. (1953). *The Problem of Multiple Comparisons. The Collected Works of John W. Tukey (1994),* **Vol. VIII** *Multiple Comparisons.* Chapman and Hall

[4] 白石高章 (2008). 多群モデルにおけるウィルコクソンの順位和に基づくノンパラメトリック同時信頼区間. 応用統計学, **37.** 125-150.

[5] 早川由宏.『Mathematica とＣ言語による統計プログラミングの基礎』. 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 年 1 月.