

多重比較法の最近の動向の研究 (Webによる多重比較法の実現)

M2004MM008 堀内 賢太郎

指導教員 松田 眞一

1 はじめに

堀内 [5] では、すべての対比較を同時に検定するテューキー (Tukey) の方法と対比較をおこなう際にステップダウン (step-down) 法 (下降手順) を用いるテューキー・ウェルシュ (Tukey-Welsch) の方法を理解し、それらをプログラム化した。

そこで今回は、最近の多重比較法について、どのような方法論が存在するのかを理解したいと考えた。調べを進めていくうちに、FDR(False Discovery Rate) を制御する多重比較法が多く見られた。この方法は、従来の第1種の過誤確率 (FWER) を制御する多重比較法とは異なり、棄却された仮説のうち、誤って棄却された帰無仮説の割合 (FDR) を制御しているところが大きな違いである。この考え方は Benjamini and Hochberg[1] によって提案されたものであり、それに付随して考え出された手法がいくつか存在していることが分かった。本論文では、BH法 (Linear step-up procedure)、adaptiveBH法、Two-stage法 (Two-stage linear step-up procedure)、BY法を取り上げた。そして、BH法、ABH法、BY法、SNK法のプログラミングをおこない、それらについてのシミュレーション評価をおこなった。

さらに、Webを用いた多重比較の実現は、perlで書かれたCGIファイルからR(統計解析ソフト)を呼び出し、関数を実行し、その結果を返すようなものを作成した。

2 FDR(False Discovery Rate) の概念と定義

同時に m 個の仮説を検定する問題を考える (m_0 は未知の真の帰無仮説の数とする)。 R は、棄却される仮説の数である (太字で確率変数を表す記号法は Benjamini and Hochberg [1] の考え方に従った)。表1は、状況を伝統的な形式にまとめたものである。特定の m 個の仮説は、前もって知られているとする。 R は、観測可能な確率変数であり、 U, V, S, T は、観測不可能な確率変数である。

| | 検定 | | 計 |
|--------|----------|----------|-----------|
| | 非有意 | 有意 | |
| 真の帰無仮説 | U | V | m_0 |
| 偽の帰無仮説 | T | S | $m - m_0$ |
| 計 | $m - R$ | R | m |

表1 m 個の帰無仮説を検定する際に犯された過誤の数

帰無仮説を誤って棄却することによって犯される過誤の割合は、確率変数 $Q = \frac{V}{V+S}$ (棄却された帰無仮説のうち、誤って棄却された仮説の割合) によって調べるこ

とができる。 $V + S = 0$ (間違った棄却の過誤を犯すことができない、つまり1つも仮説が棄却されない)の時は、 $Q = 0$ と定義する。また、 v や s の値(小文字は実現値を意味している)が分からないので、 Q は観測できない(未知の)確率変数であり、実験やデータ分析をおこなった後でさえ、このような $q = \frac{v}{v+s}$ は求めることができない値である。そこで Benjamini and Hochberg[1] は、 Q の期待値をとったものをFDRとし、 Q_e を以下のように定義した。

$$Q_e = E(Q) = E\left(\frac{V}{V+S}\right) = E\left(\frac{V}{R}\right)$$

つまり、FWER(真の帰無仮説が1つでも棄却される確率)を制御するのではなく、FDR(帰無仮説のうちの真の帰無仮説の割合)を制御することによって検出力の増加を期待できるかもしれないと考えたのである。

3 FDRを制御する多重比較法

以下で、本論文で研究した方法の手順を説明する。なお、Two-Stage法[3]とBY法[4]については省略した。

3.1 BH法

仮説 H_1, H_2, \dots, H_m はそれぞれの仮説に対応した p 値(有意確率) P_1, P_2, \dots, P_m を持つと仮定する。さらに、昇順に並べ替えられた p 値を $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$ と表し、 $P_{(i)}$ に相当する帰無仮説は $H_{(i)}$ と定義する。また、有意水準のようなもの (FDRの割合を全体としてどれぐらいに抑えたいのかという基準となる値) を q^* (通常は0.05) とする (Benjamini and Hochberg[1] 参照)。

手順1 $P_{(i)} \leq \frac{i}{m} q^*$ を満たす $i (= 1, 2, \dots, m)$ の中で最大の値 k を定める。

手順2 仮説 $H_{(i)}; i = 1, 2, \dots, k$ をすべて棄却する。また、もしそのような k が存在しないならば、どの仮説も棄却することなく終了する。

3.2 Adaptive BH法

最初の条件はBH法と同じである [2] 参照。

手順1 基準値 q^* でBH法を実行する。もし1つも仮説を棄却することができないならば、終了する。それ以外ならば手順2へ進む。

手順2 計算式 $S_i = \frac{1-P_{(i)}}{m+1-i}$ によって、 S_i を計算する ($i = 1, 2, \dots, m$)。

手順3 $i = 2$ からスタートし、 $S_i \geq S_{i-1}$ を満たしている限り i を増加させていく。 $S_i < S_{i-1}$ と初めてなっ

た時の S_i を保持する。最後まで $S_i \geq S_{i-1}$ ならば、 $i = m$ として手順 4 に進む。

手順 4 $\hat{m}_0 = \min([\frac{1}{S_i} + 1], m)$ とする ($[\]$ はガウス記号)。

手順 5 最も大きな有意確率 $P_{(m)}$ からスタートし、 $P_{(k)} \leq \frac{k}{\hat{m}_0} q^*$ を満たす k が初めて現れるまで $P_{(i)}$ と $\frac{i}{\hat{m}_0} q^*$ を段階的に比較していく。

手順 6 有意確率 $P_{(k)}$ を持つ仮説と $P_{(k)}$ より小さい有意確率を持つそれ以降すべての仮説を棄却する。つまり、仮説 $H_{(i)}; i = 1, 2, \dots, k$ すべてを棄却する。

4 Web による多重比較法

perl で書かれた CGI ファイルから R(統計解析ソフト) を呼び出し、関数を実行し、その結果を返すようなものを作成した。掲載手法は、一元配置分散分析に加え、FWER を制御する方法として、Tukey 法、Tukey-Welsch 法、Peritz 法を、FDR を制御する方法として、BH 法、ABH 法、SNK 法を掲載した。データの入力形式は BH 法、ABH 法がベクトルデータ、それ以外の 4 手法は行列データである。それぞれラジオボタンでデータの形式と解析手法をクリックし、実行させる形式となっている。また、BH 法、ABH 法についてはデータ構造も選択しないと解析できないようになっている。

5 シミュレーション

5.1 過去の比較研究

Benjamini and Hochberg[1] では、BH 法、Bonferroni の方法、Hochberg の方法の 3 手法を取り上げ、仮説間が独立な場合の検出力のシミュレーションをおこなっていた。結果は、真の仮説と偽の仮説がどんな割合で混じっていても検出力は BH 法がすべて勝っていた。つまり、FWER を制御する多重比較法よりも FDR を制御する多重比較法の方が検出力が高いことを示している。

また、FDR を制御する多重比較法のシミュレーション比較として、Benjamini and Hochberg[2] では、BH 法と ABH 法が比較されていた。仮説間が独立な場合、検出力が BH 法よりも高い手法として提案されていた ABH 法が、極端な仮説の配置を除いて勝っていた。

さらに、一番多くの手法をシミュレーション比較している Benjamini, Krieger and Yekutieli[3] では、本論文で取り上げていないような手法がいくつか提案されていて、様々な状況下でのシミュレーションがなされていた。この論文では、真の仮説と偽の仮説の割合、仮説間の相関(正の相関)、仮説数をいろいろと変化させてシミュレーションをおこなっていた。結果として、仮説間が独立な場合、ABH 法は、Two-Stage 法よりも検出力が高いこと。また、このような状況下では最も高い検出力が得られる手法として、Resampling をおこなう方法が示されていた。

5.2 シミュレーション

私は、以下の 4 種類の方法において、さまざまな状況下でのシミュレーション評価をおこなっている論文が見当

たらなかったので、取り組んだ。また、Two-Stage 法については、特殊な場合でのシミュレーション評価が書かれている論文 [3] 内の正の相関がある場合のシミュレーション結果を見てみると、ABH 法のほうが Two-Stage 法よりも高い値をとっていたので、同じ状況で使用できる方法の中の代表として ABH 法を選び、Two-Stage 法についてはシミュレーション評価をおこなっていない。

(シミュレーションの目的)

- BH 法、ABH 法、BY 法、SNK 法が FDR を 0.05 以下に保っているのかを確認する。

(シミュレーションパターン)

- 1 <仮説間が独立な場合>
(パターン 1)
- 2 <仮説間がすべて正の相関である場合>
(パターン 2,3,4,8,9)
- 3 <仮説間に負の相関がある場合>
(パターン 5,6,7,10)

(シミュレーション方法)

正規乱数を用いて、1 標本検定をおこなった。ここでは、真の帰無仮説を平均 0、偽の帰無仮説を平均 1 とし、シミュレーション回数は 10000 回である。

パターン 1 については、それぞれ独立に P 値を生成し、仮説間に正の相関構造を持たせる場合、パターン 2 はコレスキー分解を用いて生成し、パターン 3,4 については移動平均を用いて、仮説間に相関構造を持たせた。さらに、負の相関構造を持たせる場合、パターン 5,6 については、移動平均を取ったものを反転させることによって生成した。

また、パターン 7 については対比較の設定で、2 標本検定をおこなった。真の帰無仮説を平均が等しい、偽の帰無仮説を平均の差が 1 であると、 P 値を生成した。

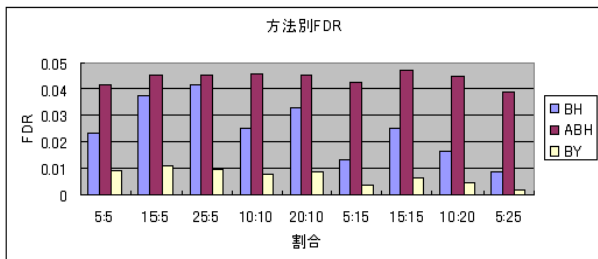
(シミュレーションの正当性)

BH 法については、特殊な場合でのシミュレーション評価が書かれている論文 [3] のデータを、ABH 法については、ABH 法が提案されていた論文 [2] のデータを用いてシミュレーションの正当性を確認した。それぞれの手法における FDR 値は、論文のものと同ほぼ同じ FDR 値を取っていた。

BY 法については、手順からも分かる通り $q^{***} = q^* / \sum_{j=1}^m (\frac{1}{j})$ とし、 q^* の代わりに q^{***} を用いて、BH 法をおこなっている点に注目してほしい。これは、仮説数に依存して、基準値である q^* を 0.05 からそれよりも小さな値になるように調整しているので、保守的な方法であることに間違いは無い。従って、BY 法はかなり厳しく FDR を制御している方法なのである。また、(佐藤ら [6]) でも BY 法はかなり保守的な方法であると書かれていた。

6 シミュレーションによる評価と考察

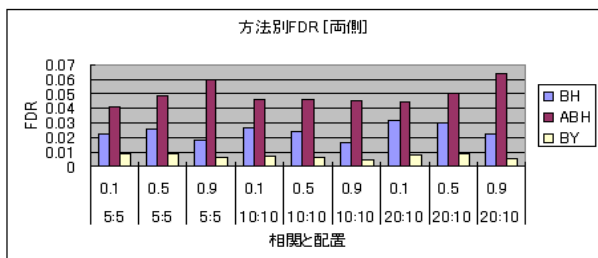
6.1 <パターン 1(独立)>



● 真の仮説と偽の仮説の配置を 5:5,15:5,25:5,10:10, 20:10, 5:15,15:15,10:20,5:25 の 9 パターンを考えシミュレーションをおこなった。

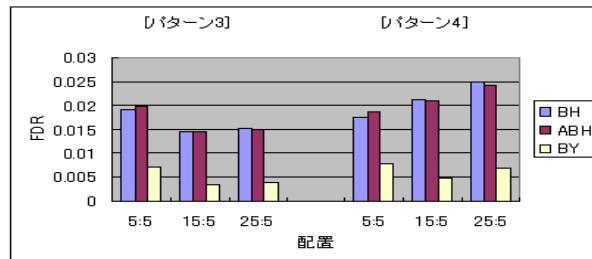
すべての場合において ABH 法の FDR 値がその他の方法の FDR 値よりも勝っていることが分かった。また、データ間に相関がある場合に用いると記されていた BY 法は、相関がない場合は FDR 値はかなり低く、かなり保守的な方法であることが分かった。さらに、BH 法において、真の仮説の数を 5 に固定して表を見た場合、FDR 値は減少しているが、逆に偽の仮説の数を 5 に固定して表を見た場合、FDR 値は増加していた。それに比べて、ABH 法の FDR 値においては、いずれの場合も安定した FDR 値を取っていた。

6.2 <パターン 2,3,4(正の相関)>



● 変数間のすべての相関係数が $\text{cor}=0.1, 0.5, 0.9$ の 3 パターンにおいて、独立な場合と同様の 9 パターンを考えシミュレーションをおこなった。表は、5:5,10:10,20:10 のパターン別 FDR 値である。

この結果から言えることは、BH 法と BY 法は、FDR 値が 0.05 以下に保たれていて、ABH 法においては、真の仮説が総仮説数の 1/2 以上存在する場合で、仮説間の相関が 0.5 以上のものはほとんど FDR 値が 0.05 を上回っていた。それ以外の場合では、ほぼ全部 ABH 法が一番高い FDR 値を取ることが分かった。従って、データ間に相関関係がある場合でも、BH 法を手法として用いることは間違いではないことが分かった。また、仮説間に正の相関関係がある場合、ABH 法を適用することは正しくない。さらに、BY 法は仮説間に相関関係がある場合でも、かなり保守的な方法論であるといえる。

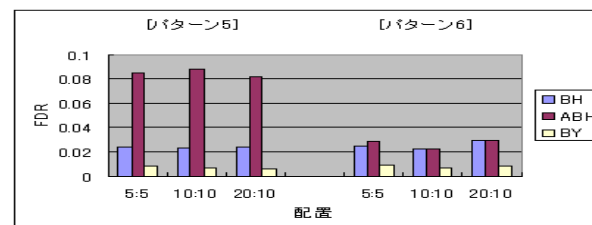


● 隣り合う仮説間には強い正の相関があり、離れている仮説間には弱い正の相関がある場合 (パターン 3,4)。

パターン 3 において、BH 法と ABH 法では、仮説の配置 (真と偽) が異なっても FDR 値は 2 手法とも安定して保たれていた。以前の結果から、ABH 法は仮説間に正の相関関係 ($\text{cor}=0.5, 0.9$) がある場合、FDR 値は 0.05 を超えていたが、今回は BH 法とほぼ同じ値を取り、FDR 値はいずれの場合も 0.05 以下に保たれていた。

パターン 4 においては、真の仮説が多いと高い FDR 値を取っており、偽の仮説が多いと低い FDR 値を取っていた。

6.3 <パターン 5,6,7(負の相関)>



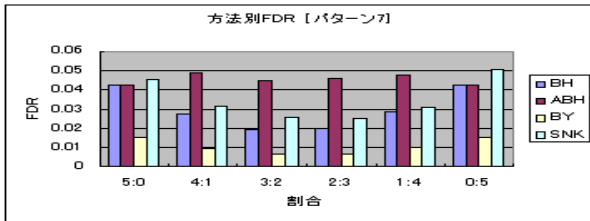
この場合も独立な場合と同様の 9 パターンを考え、シミュレーションをおこなった。表は、5:5,10:10,20:10 のパターン別 FDR 値である。

● 仮説間に正の相関と負の相関が交互に現れる相関構造がある場合。つまり、真の仮説の中身がすべて正の相関を持っていて、真の仮説と偽の仮説の間で負の相関を持つような構造の場合 (上図:パターン 5)。

ABH 法はいずれの場合も、FDR 値が 0.05 を超えていた。それに対して、BH 法は ABH 法よりも安定した FDR 値を得ることができて、尚且つ FDR 値が 0.05 以下に保たれていた。さらに、BH 法は仮説の配置 (真と偽) が異なっても FDR 値を安定して保っていた。

● 仮説間で適度に正と負が混じり合ったような相関構造を持つ場合 (上図:パターン 6)。

いずれの手法も FDR 値が 0.05 以下に保たれていて、BH 法と ABH 法は、ほとんど同じような FDR 値を取っていた。しかし、真の仮説の数が 5 の場合のみ、BH 法と ABH 法の FDR 値に差が見られた。値としては、ABH 法が BH 法よりも高い FDR 値を取っていた。また、いずれの手法も仮説数が増加するにつれて、FDR 値が減少していった。

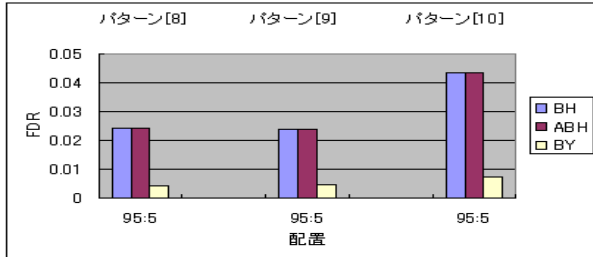


● 対比較をおこなった場合 (パターン 7)。

全部で 5 群あるうち、平均 0:平均 1 が 5:0,4:1,3:2,2:3, 1:4,0:5 の計 6 パターンを考えている。全体的に正の相関や負の相関が混じっていて、仮説間の相関構造が 0.5 前後であっても、いずれの手法で FDR 値が 0.05 以下に保たれていることが分かった。また、ABH 法の FDR 値は、いずれの場合も安定しており、非常に高い FDR 値を取っていることが分かった。さらに、BH 法において、仮説がすべて真の場合 (5:0,0:5) は FDR 値は高く、仮説に真と偽がほぼ均等に配置されている場合 (3:2,2:3) は、他よりも厳しめに FDR 値を守っていることが分かった。

また、SNK 法は、平均値がどのような配置でも FDR 値を 0.05 以下に保っていることは既に証明済みであるが、確認のためシミュレーションをおこなった。その結果、FDR 値は誤差を除けば 0.05 以下に保っていることが確認できた。

6.4 <パターン 8,9,10 >



● 偽の仮説の数を固定して、真の仮説の数を増やしていった場合 (パターン 8,9 は正の相関を持つもので、パターン 10 は負の相関を持つものである)

最大仮説数 100 まではパターン 3,4,6 と同じ結果となり、いずれも FDR 値は 0.05 以下に保たれていた。

7 まとめ

○ BH 法は仮説間にどのような相関関係があっても、安定した検定結果がえられる手法である。従って、仮説間が独立ではなく、相関関係 (正と負) があったとしても、ある程度信頼性のある結果を得ることができる。

○ ABH 法はある特殊な状況 (正の相関が強く、仮説に真の仮説が多く含んでいる場合や仮説間に正の相関と負の相関が交互に現れる場合) を除けば、BH 法と同様な安定した FDR 値を得ることができる手法である。さらに、ABH 法は仮説間に相関がない場合や低い場合、対比較を

おこなう場合においては、BH 法に勝る。

○ BY 法は BH 法と同様、仮説間にどのような相関構造があっても、用いることは可能である。しかし、FDR を制御する多重比較法の立場から見るとかなり保守的な手法である。

8 おわりに

本論文では、最近の多重比較法の動向を調査、研究することを目的とした。その中で、FDR を制御する多重比較法を発見し、それに付随した方法も含めた形でシミュレーション評価をすることができた。そこでは、BH 法の安定性、ABH 法が強い状況、などを確認することができた。

また、Web による多重比較法の実現として、従来の多重比較法に加え、FDR を制御する多重比較法も掲載することができた。

今後の課題として、どのような場合にどのような手法を用いれば適切なのかを明確にすることである。今のところ、仮説間に相関構造があったとしても、BH 法を適用することは間違いではないと考えられる。

謝辞

本論文を作成するにあたり熱心に御指導頂きました、松田眞一先生には深く感謝いたします。また、その他協力して頂いた田中豊先生、木村美善先生、富田誠先生に深く感謝いたします。

参考文献

- [1] Y.Benjamini and Y.Hochberg: "Controlling the false discovery rate: a practical and powerful approach to multiple testing", JRSS seriesB, no.57, 289-300(1995).
- [2] Y.Benjamini and Y.Hochberg: "On the adaptive control of the False Discovery Rate in multiple testing with independent statistics" Journal of Educational and Behavioral Statistics, vol.25, no.1, 60-83(2000).
- [3] Y.Benjamini, A.M.Krieger and D.Yekutieli: "Adaptive Linear Step-up Procedures that control the False Discovery Rate", (<http://www.math.tau.ac.il/~ybenja/MyPapers/bkymarch9.pdf>)(2005).
- [4] Y.Benjamini and D.Yekutieli: "The control of the false discovery rate in multiple testing under dependency", AS.29,no.4,1165-1188(2001).
- [5] 堀内賢太郎: "R による多重比較法の研究", 南山大学 数理情報学部 数理科学科 卒業研究 (2004).
- [6] 佐藤亜香里・角谷伸一・田崎武信: "FDR コンセプトと FDR 法についての考察", 塩野義製薬株式会社解析センター (2004).