分割表における多重比較法とその評価

M2004MM040 棚瀬 貴紀
指導教員　松田 眞一

1　はじめに

分割表（クロス表）は統計的手法の中でもごく基本的なツールであるといえる。その用途は幅広く、社会経済からマーケティング、医薬にまで及んでいる。

しかし、一般に使われている分割表の検定である Pearson の χ^2 検定や Fisher の正確確率検定は、縦と横に関連があるかどうかを調べるに留まっている。さらにどのカテゴリー間に関連があるかどうかを調べるには、有意水準の多重性を考慮する、多重比較法の概念が必要となってくる。最近では対数線形モデルを用いた方法がとられているが、それは分割表そのものからのアプローチではない。現在提唱されている多重比較法が得意としているパターンを見出し、望まれる適用を提案したい。

2　本論における分割表の設定

要因 A が a 個の尺度、要因 B が b 個の尺度があるように分類した 2 元分割表を仮定する。分割表のセル (i,j) にて観測された度数を n_{ij} とし、総数を $\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij} = N$、第 i 行や第 j 列に関しての度数の和を R_i, C_j とする。また、各セルの出現同時確率を p_{ij} とする。$\sum_i p_{ij} = p_{i.}$、$\sum_j p_{ij} = p_{.j}$、$\sum_i\sum_j p_{ij} = p_{..} = 1$ である (表 1 参照)。

表 1　$a \times b$ 分割表

	B_1	$\cdots$	B_j	$\cdots$	B_b	計
A_1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1b}	R_1
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{ib}	R_i
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_a	n_{a1}	$\cdots$	n_{aj}	$\cdots$	n_{ab}	R_a
計	C_1	$\cdots$	C_j	$\cdots$	C_b	N

3　広津の累積 χ^2 法

周辺和で条件付けした分布は、一般に多項一般超幾何分布であり、多項分布の一様性の仮説の下では多項超幾何分布となる。

$a \times b$ 分割表の行と列に関する確率変数ベクトルを

$$\boldsymbol{r}' = N^{-\frac{1}{2}}(\sqrt{R_1}, ..., \sqrt{R_i}, ..., \sqrt{R_a}) \tag{1}$$

$$\boldsymbol{c}' = N^{-\frac{1}{2}}(\sqrt{C_1}, ..., \sqrt{C_j}, ..., \sqrt{C_b}) \tag{2}$$

とおく。R'、C' は $\boldsymbol{rr}' + RR' = I_a$、$\boldsymbol{cc}' + CC' = I_b$ となるような $(a-1)\times a$、$(b-1)\times b$ 行列とする。

$$\boldsymbol{y}' = \{y_{11}, y_{12}, ..., y_{1b}, y_{21}, ..., y_{2b}, ..., y_{a1}, ..., y_{ab}\} \tag{3}$$

ただし $y_{ij} = n_{ij}/\sqrt{R_i C_j/N}$

とおくと、この $\boldsymbol{y}$ は漸近的に平均 $\sqrt{N}(\boldsymbol{r}\otimes\boldsymbol{c})$、分散 $(RR')\otimes(CC')$ の正規分布に従う。さらに$\boldsymbol{r}$、$\boldsymbol{c}$それぞれに直交する任意の単位ベクトルを$\boldsymbol{r}_\perp$、$\boldsymbol{c}_\perp$とすると、$(\boldsymbol{r'}_\perp \otimes \boldsymbol{c'}_\perp)\boldsymbol{y}$ は標準正規分布に従うことが言える。

3.1　適合度 χ^2

分割表の検定において最も単純なものは、明示的な対立仮説を与えずに帰無仮説を検定する適合度 χ^2 検定である。$(R'\otimes C')(RR'\otimes CC')(R\otimes C) = I_{(a-1)\times(b-1)}$ より、確率ベクトル $(R'\otimes C')\boldsymbol{y}$ の $(a-1)(b-1)$ 個の確率変数が互いに独立に標準正規分布に従うことが分かる。さらに、これらの 2 乗和 $\| (R'\otimes C')\boldsymbol{y} \|^2$ は $\chi^2_{(a-1)(b-1)}$ に従い、Pearson の χ^2 検定と同質である。

またこれは分割表全体に関する統計量である。ある 2 行間 $(i \neq i')$ のみに限定して述べれば、仮説は

$$H_0(i,i') : p_{ij}/p_{i.} - p_{i'j}/p_{i'.} = 0 \quad (j = 1, ..., b) \tag{4}$$

となる。これに対し、行間の 2 乗距離の統計量は以下のようにかける (Hirotsu[2] 参照)。

$$\chi^2(i;i') = \| \{\boldsymbol{r}'(i;i')\otimes C'\}\boldsymbol{y} \|^2 \tag{5}$$

ただし

$$\boldsymbol{r}'(i;i')$$

$$= \left(\frac{1}{R_i} + \frac{1}{R_{i'}}\right)^{-\frac{1}{2}}\left(0, ..., 0, \frac{1}{\sqrt{R_i}}, 0, ..., 0, -\frac{1}{\sqrt{R_{i'}}}, 0, ..., 0\right)$$

3.2　累積 χ^2 法

前節は列カテゴリーに順序がないことを想定していたが、カテゴリーに自然な順序があると統計量の計算は非対称となる。自然な順序とは、例えばガンの重症度を要因とすれば“ 軽度 ”“ 中程度 ”“ 重度 ”というような順序付き水準を示す。この節では自然な順序があることを想定し、全ての $i \neq i'$ に対して傾向のある対立仮説

$$H_1(i,i') : p_{i1}/p_{i'1} \geq ... \geq p_{ib}/p_{i'b} \quad \text{or} \quad p_{i1}/p_{i'1} \leq ... \leq p_{ib}/p_{i'b} \tag{6}$$

を与え、この仮説検定における統計量をこの節で考える。

分割表全体に対して列 j と列 $(j+1)$ の間に仕切りを入れてまとめた $a\times 2$ 分割表に対して χ^2 統計量 $\chi^2_{(j)}$ を考える。この統計量の累積和は

$$\chi^{*2} = \sum_{j=1}^{b-1}\chi^2_{(j)} = \| (R'\otimes C^{*\prime})\boldsymbol{y} \|^2 \tag{7}$$

とかける。ただし $C^{*\prime}$ の第 j 行 $\boldsymbol{c}^{*\prime}_j$ $(j = 1, ..., b-1)$ は

$$\boldsymbol{c}^{*\prime}_j = \left(\frac{1}{S_j} + \frac{1}{\bar{S}_j}\right)^{-\frac{1}{2}}\left(\frac{\sqrt{C_1}}{S_j}, ..., \frac{\sqrt{C_j}}{S_j}, \frac{-\sqrt{C_{j+1}}}{\bar{S}_j}, ..., \frac{-\sqrt{C_b}}{\bar{S}_j}\right)$$

$$(S_j = \textstyle\sum_{l=1}^{j} C_l,\ \ \bar{S}_j = N - S_j\ \ (j = 1, ..., b-1))$$

である。これは χ^2 分布の定数倍 $d\chi^2_\nu$ でよく近似される。d, ν については

$$d = 1 + \frac{2}{b-1}\sum_{j=1}^{b-2}\left(\frac{\sum_{k=1}^{j}\lambda_k}{\lambda_{j+1}}\right) \tag{8}$$

$$\nu = (a-1)(b-1)/d \tag{9}$$

と求まる。

3.3 Scheffé 法

行を群とみなし、一般性を失うことなくこれらを 2 群 $I = (1, ..., u)$、$I' = (u+1, ..., u+v)$ に分ける。このとき、目的の統計量は以下のようになる。

$$\chi^{*2}(I, I') = \| \{\boldsymbol{r}'(I; I') \otimes C^{*\prime}\}\boldsymbol{y} \|^2 \tag{10}$$

ただし $\sum_{i \in I} R_i = T_I$、$\sum_{i \in I'} R_i = T_{I'}$ で

$$\boldsymbol{r}'(I; I') = \left(\frac{1}{T_I} + \frac{1}{T_{I'}}\right)^{-\frac{1}{2}}$$

$$\times \left\{\frac{\sqrt{R_1}}{T_I}, ..., \frac{\sqrt{R_u}}{T_I}, \frac{-\sqrt{R_{u+1}}}{T_{I'}}, ..., \frac{-\sqrt{R_{u+v}}}{T_{I'}}, 0, ..., 0\right\}$$

そこで Hirotsu[2](他、広津 [3]) は、$\| \{\boldsymbol{r}'(I; I') \otimes C^{*\prime}\}\boldsymbol{y} \|^2$ の最大値が Wishart 分布 $W(C^{*\prime}C^*, a-1)$ に従う行列の最大固有根の分布に、H_0 の下で漸近的に従うことを証明した。

しかし、この分布から棄却限界値を定める計算が非常に複雑であることから、会田・広津 [1] が求めた数表 (ただし $b = 3, 4$) を用いるか、$\gamma\chi^2_{a-1}$(γは $C^{*\prime}C^*$の最大固有根) にて近似を行うことが出来る。

また、2 群それぞれを行間として当てはめた統計量は $\chi^{*2}(i; i') = \| \{\boldsymbol{r}'(i; i') \otimes C^{*\prime}\}\boldsymbol{y} \|^2$ となる。これもやはり (7) の成分であるが、群間を考える (10) と比べて相当保守的な統計量であると考えられる。

3.4 Tukey 法

基本統計量は $\chi^{*2}(i; i')$ である。特に a があまり大きくないときに以下の不等式の下限が十分によい近似を与えるとしている (広津 [3])。

$$\binom{a}{2}\Pr(d\chi^2_\nu > c) - \sum \Pr\{\chi^{*2}(i_1; i_2) > c, \chi^{*2}(i_3; i_4) > c\}$$
$$\leq \Pr\{\max_{i \neq i'} \chi^{*2}(i; i') > c\} \leq \binom{a}{2}\Pr(d\chi^2_\nu > c) \tag{11}$$

ただし、$\sum$ は $(i_1, i_2) \neq (i_3, i_4)$ に関してすべての和をとることを意味する。

しかしこの有意確率をシミュレーションにて試した結果、$a = 3$ のときに上限の値が 1 を超えてしまうパターンが、$a \geq 4$ のとき下限の値が負になるかつ上限の値が 1 を超えるパターンが発見された。ここで説明される Tukey 法において実用可能なのは $a = 3$ のみであり、それ以上の行数では近似が悪く用いることは出来ない。任意の対比較を行いたいときは、以降で述べる松田の方法や広津・松田の方法を用いるべきである。

4 松田の χ^2 法

$a \times b$ 分割表 $(a \geq 3)$ に対して、松田 [4] は閉検定手順 (Marcus *et al.*[5] 参照) を用いた方法を提案している。その方法を以下に述べる。

まず、次のような仮設の族を考えている。

$$S = \{p_{ij}/p_{i.} = p_{i'j}/p_{i'.}\,(1 \leq j \leq b)\,|\,1 \leq i \leq i' \leq a\}$$

すべての仮説が成り立つ場合は、独立性の帰無仮説に一致する。さらに仮説を拡張し、いくつかの行の集合でのみ構成される仮説の族 S' を

$$S' = \{p_{i_1 j}/p_{i_1 .} = \cdots = p_{i_k j}/p_{i_k .}\,(1 \leq j \leq b)$$
$$|\, i_1, ..., i_k \subseteq \{1, 2, ..., a\}\}$$

とする。k は仮説の大きさと呼ばれる。この S' に対して以下のように閉検定手順が適用される。

手順 1　有意水準 α を定める。

手順 2　分割表全体に対して有意水準 α で独立性の検定を行う。棄却されれば $k = a-1$ とおいて次へ進む。保留されれば S' の全ての仮説を保留して終了する。

手順 3　S' でまだ保留となっていない大きさ k の全ての仮説について、対応する行のみを取り出した部分分割表に対する有意水準 α_k の独立性の検定を行う。ただし、

$$\alpha_{a-1} = \alpha, \alpha_k = 1 - (1-\alpha)^{k/a}\ (k = 2, ..., a-2)$$

である。

手順 4　大きさ k のすべての仮説が保留されている場合、または $k = 2$ 場合は終了する。

手順 5　S' 内で保留された仮説を含む (行の添え字の集合としては含まれる) 大きさ $k-1$ の仮説を全て保留する。大きさ k の仮説のうち、一つでも棄却されれば $k-1$ を新しく k とおき、**手順 3** に戻ってこの手順を繰り返す。

手順 2、**手順 3** で用いられる独立性の検定は、Pearson の χ^2 検定と Fisher の正確確率検定のどちらでも良いが、松田 [4] はどちらか一方のみの使用を推奨している。

5 提案する方法

基本は閉検定手順を用いている点では、松田の方法と同質である。松田の方法は独立性の検定にて“ Pearson の χ^2 検定と Fisher の正確確率検定のどちらか ”としていたが、提案する方法は、この独立性の検定にて累積 χ^2 検定を用いる点のみで異なる。順序を仮定した対立仮説の下で、累積 χ^2 法が高い検出力を持つことが知られている (会田・広津 [1] など)。このことから、この方法も同様の性質を持つことが期待される。

本論文では、この方法を広津・松田の方法とよぶことにする。なお、この方法が所与の有意水準を守ることは松田の方法と同様に示される。

6 検出力シミュレーション

本研究の最大の目的である任意の行間の対比較に対する、検出力シミュレーションを行った。

行数 $a = 3, 4, 5, 6$、列数 $b = 3$ に固定する。繰り返し数は 10000 回とした。検証する方法は以下の 5 つとする。

(i) 松田の方法
(ii) 統計量 $\chi^{*2}(i, i')$ の $W(C^{*\prime}C^{*}, a-1)$ による検定
(iii) 広津・松田の方法
(iv) 累積 χ^2 法
(v) Pearson の χ^2 検定

(i)〜(iii) は多重比較法で、(iv)〜(v) は独立性の検定である。

6.1 傾向のある対立仮説の場合

列に順序のある対立仮説を仮定する。すなわち、対立仮説として (6) を仮定する。この仮説が真であるとき、度数は左寄りか右寄りに出現する。

シミュレーションを行う方法として、各セルの発生確率 p_{ij} を適当に定め、乱数により度数を決定する。まず標準正規分布の分布関数の値が 1/3、2/3 になる点を U_1、U_2 とする。ここで $\boldsymbol{\mu} = (\mu_1, ..., \mu_i, ..., \mu_a)$ を適当に定め、$N(\mu_i, 1^2)$ の密度関数の区間 $(-\infty, U_1)$、(U_1, U_2)、(U_2, ∞) での累積確率をそれぞれ q_{i1}、q_{i2}、q_{i3} とする。さらに $p_{\cdot\cdot} = 1$ に固定するため、$p_{ij} = q_{ij}/a$ とする。$\boldsymbol{\mu}$ の設定は、区間 [0, 0.5] と区間 [0, 0.7] を行数だけ等分するパターン A、B の二つを用意した。シミュレーション結果を表 2〜5 に載せる (下線は各区分の最大値)。

表 2　傾向のある対立仮説の下での検出力 (3 × 3 分割表)

パターン	$N = 150$(上段)、$N = 300$(下段)				
	多重比較法			独立性の検定	
	(i)	(ii)	(iii)	(iv)	(v)
A	0.3786	0.4275	0.4528	0.4700	0.3790
B	0.6637	0.7267	0.7566	0.7655	0.6656
A	0.6993	0.7545	0.7873	0.7910	0.7003
B	0.9508	0.9704	0.9754	0.9755	0.9501

表 3　傾向のある対立仮説の下での検出力 (4 × 3 分割表)

パターン	$N = 200$(上段)、$N = 400$(下段)				
	多重比較法			独立性の検定	
	(i)	(ii)	(iii)	(iv)	(v)
A	0.3291	0.3482	0.4312	0.4508	0.3483
B	0.6234	0.6465	0.7318	0.7507	0.6487
A	0.6538	0.6813	0.7704	0.7807	0.6785
B	0.9415	0.9474	0.9724	0.9751	0.9467

表 4　傾向のある対立仮説の下での検出力 (5 × 3 分割表)

パターン	$N = 250$(上段)、$N = 500$(下段)				
	多重比較法			独立性の検定	
	(i)	(ii)	(iii)	(iv)	(v)
A	0.2866	0.2687	0.4048	0.4536	0.3533
B	0.5824	0.5778	0.7291	0.7660	0.6590
A	0.6277	0.6160	0.7655	0.7872	0.6808
B	0.9351	0.9297	0.9701	0.9806	0.9547

表 5　傾向のある対立仮説の下での検出力 (6 × 3 分割表)

パターン	$N = 300$(上段)、$N = 600$(下段)				
	多重比較法			独立性の検定	
	(i)	(ii)	(iii)	(iv)	(v)
A	0.2425	0.2088	0.3800	0.4727	0.3533
B	0.5421	0.5106	0.7124	0.7885	0.6704
A	0.5767	0.5482	0.7415	0.8155	0.7079
B	0.9324	0.9145	0.9724	0.9841	0.9658

多重比較法に関しては (iii) 広津・松田の方法が、独立性の検定に関しては (iv) 累積 χ^2 法が全面的に最も高い検出力を得る結果となった。(i) と (ii) の比較では、$a = 3, 4$ のとき (ii) が勝っているが $a \geq 5$ で (i) が逆転する結果となった。なお、閉検定手順を用いている (i)(iii) は最初に全体の独立性の検定を行う性質から、それらの検出力は独立性の検定 (v)(iv) の検出力を上回ることはできない (一部で関係が逆になっているものがあるが、シミュレーション誤差であろう)。

以上をまとめると、単調な傾向のある対立仮説では Pearson の χ^2 よりも累積 χ^2 を考える統計量の方が検出力は高いことが言える。累積 χ^2 は複数方向への射影成分を用いるため、単調に増加するベクトルに逐一射影することから今回のような検出力を得ることができた。対して Pearson の χ^2 は全体同一方向から射影をとるために、累積 χ^2 と比較して検出力は落ちる。

6.2 傾向のない対立仮説の場合

累積 χ^2 法関連は列方向に自然な順序を仮定しているため、単調な傾向を仮定しない状況では検出力が落ちることが想像される。例えば、以下の対立仮説は非単調な傾向を仮定する対立仮説の一つである。

$$H_1(i, i') : p_{i1}/p_{i'1} \geq ... \leq p_{ib}/p_{i'b} \quad \text{or} \quad p_{i1}/p_{i'1} \leq ... \geq p_{ib}/p_{i'b} \tag{12}$$

もしこの仮説が成り立っているのであれば、度数は中央寄りか左右寄りに出現する。

前節のシミュレーションの設定は正規分布の平均をずらすことで対立仮説を設定したが、今回は平均を固定し分散をずらすことで対立仮説 (12) を設定する。前節と異なる点は $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_i, ..., \sigma_a)$ を適当に定め、$N(0, \sigma_i^2)$ の密度関数の区間を考えることのみである。$\boldsymbol{\sigma}$ の設定は、

区間 [1, 2] と区間 [0.5, 1] を行数だけ等分するパターン A、B の二つを用意した。

シミュレーション結果を表 6〜9 に載せる。

表 6　傾向のない対立仮説の下での検出力 (3 × 3 分割表)

パターン	$N=150$(上段)、$N=300$(下段)				
	多重比較法			独立性の検定	
	(i)	(ii)	(iii)	(iv)	(v)
A	0.2854	0.0539	0.0825	0.0932	0.2981
B	0.6048	0.2732	0.3314	0.3566	0.5976
A	0.5532	0.0941	0.1474	0.1532	0.5678
B	0.9088	0.6733	0.7337	0.7467	0.9179

表 7　傾向のない対立仮説の下での検出力 (4 × 3 分割表)

パターン	$N=200$(上段)、$N=400$(下段)				
	多重比較法			独立性の検定	
	(i)	(ii)	(iii)	(iv)	(v)
A	0.2446	0.0330	0.0692	0.0915	0.2759
B	0.5522	0.1498	0.2623	0.3239	0.5798
A	0.5099	0.0483	0.1161	0.1433	0.5315
B	0.8940	0.5114	0.6643	0.7041	0.9093

表 8　傾向のない対立仮説の下での検出力 (5 × 3 分割表)

パターン	$N=250$(上段)、$N=500$(下段)				
	多重比較法			独立性の検定	
	(i)	(ii)	(iii)	(iv)	(v)
A	0.2030	0.0201	0.0702	0.0930	0.2714
B	0.5042	0.0877	0.2202	0.3187	0.5701
A	0.4641	0.0259	0.0951	0.1509	0.5399
B	0.8894	0.3625	0.6178	0.7054	0.9078

表 9　傾向のない対立仮説の下での検出力 (6 × 3 分割表)

パターン	$N=300$(上段)、$N=600$(下段)				
	多重比較法			独立性の検定	
	(i)	(ii)	(iii)	(iv)	(v)
A	0.1670	0.0107	0.0520	0.0867	0.2595
B	0.4608	0.0538	0.1756	0.3177	0.5858
A	0.4267	0.0176	0.0747	0.1454	0.5398
B	0.8654	0.2538	0.5552	0.7068	0.9241

シミュレーション結果をまとめる。多重比較法に関しては (i) 松田の方法が、独立性の検定に関しては (v)Pearson の χ^2 検定が全面的に最も高い検出力を得る結果となった。また、(iii) 広津・松田の方法は、(ii) 統計量 $\chi^{*2}(i,i')$ の $W(C^{*\prime}C^{*},a-1)$ による検定を全面的に上回っている。以上からまとめると、対立仮説 (12) では累積 χ^2 よりも Pearson の χ^2 を考える統計量の方が検出力は高いことが言える。累積 χ^2 は複数方向への射影成分を用いることから、ズレのない部分 (今回は第 1,3 列) の分だけかえって統計量に損な部分が出てしまう。対して Pearson の χ^2 は全体同一方向から射影をとるために、累積 χ^2 と比較して検出力は上がる。

また、任意の対比較においては対立仮説の設定に限定されず、(ii) 統計量 $\chi^{*2}(i,i')$ の $W(C^{*\prime}C^{*},a-1)$ による検定を用いることはやはり避けるべきであろう。

7　おわりに

本研究では分割表の多重比較法を扱い、対立仮説に順序を仮定する場合は広津・松田の方法、仮定しない場合は松田の方法が推奨される結果となった。

今後の課題としては、3 点が挙げられる。1 つ目に、3 節で述べた Tukey 法の整理がある。2 つ目に、予備検定を含めた統計手法一連の確立がある。今回の検出力シミュレーションにて方法各々の得意・不得意が明らかにされたが、実際の解析にはどの方法を用いるべきかを判断しなければならない。3 つ目に、対照群を設定する Dunnet 法と順序水準付きの Williams 法の分割表における提案がある。

謝辞

本研究に熱心にご指導をして頂いた松田眞一助教授に深く感謝申し上げます．累積 χ^2 関連の創始である明星大学の広津千尋先生は、私の拙い質問にも丁寧にお答え下さいました．この場を借りて心より感謝申し上げます．学内の田中豊先生、木村美善先生、富田誠先生からは日頃から統計学の研鑽の場を頂きました．ここに深く感謝いたします．

参考文献

[1] 会田雅人, 広津千尋: 順序制約下での多項分布比較の一方法と数表, 応用統計学, **12**, 3, 101-110 (1983).

[2] Hirotsu, C.: Defining the pattern of association in two-way contingency tables, *Biometrika*, **70**, 3, 579-589 (1983).

[3] 広津千尋: 実験データの解析—分散分析を超えて—, 共立出版 (1992).

[4] 松田眞一: 名義尺度の分割表に対する多重比較法, “南山大学紀要「アカデミア」数理情報編”, **4**, 29-37 (2004).

[5] Marcus, R., Peritz, E., Gabriel, K. R.: On closed testing procedures with special reference to ordered analysis of variance, *Biometrika*, **63**, 3, 655-660 (1976).