原 著

最大対比法を活用するための SAS/IML プログラム SAS/IML Program for Computing Probabilities Related to Maximum Contrast Methods

西山 智^{*1,*2}・柳原 宏和^{*3}・吉村 功^{*2}

Hiroshi Nishiyama^{*1,*2}, Hirokazu Yanagihara^{*3} and Isao Yoshimura^{*2}

^{*1}アベンティス ファーマ株式会社 生物統計・データマネジメント部

^{*2}東京理科大学大学院 工学研究科

^{*3}筑波大学 社会工学系

^{*1}Biometrics & Data Management, Aventis Pharma Ltd.

^{*2}Graduate School of Engineering, Tokyo University of Science

^{*3}Inst. of Policy & Planning Sciences, University of Tsukuba

e-mail: hiroshi.nishiyamaA@aventis.com

Many maximum contrast methods have been used to evaluate the dose-response relationship in clinical trials and in toxicological trials. However, the calculation method for p -value and power of such methods in practical form have not been proposed due to the complex property of integration in the multivariate t -distribution with singular correlation matrix. Addressing this issue, this paper presents SAS/IML programs for computing p -value, power and sample size necessary in the application of maximum contrast methods. It was confirmed that the speed of calculation in the proposed programs was far superior to the naive Monte-Carlo simulation keeping the precision as well.

Key words: Maximum contrasts method, Multivariate t -distribution, p -value calculation, Sample size calculation, SAS/IML.

1. は じ め に

より正確な表現は次章で述べるが、最大対比法 (maximum contrast method) とは、一組の線形仮説のそれぞれに対応する対比を用意し、それをスチューデント化した統計量、すなわち「対比統計量」の最大値を検定統計量とする検定法である。最大対比法に属する検定法で有名なものには、ダネット検定 (Dunnett, 1955) やチューキー検定 (Hochberg and Tamhane, 1987)^{*}があり、臨床試験や毒性試験に頻用されている。これに関連して Stewart and Ruberg (2000) は、臨床試験や毒性試験での用量反応性の評価にこれら以外の最大対比法を用いると良い場合があることを指摘している。

彼らが主張するように、ダネット検定などの古典的な最大対比法でないものが実際問題では有

^{*}文献表にある Tukey の原論文は現在入手できないので紹介文献を引用する。

用であると感じている実際家は少なくない。しかしそれを有効に利用した例は稀である。その原因は、任意の対比の組に対する、 p 値、検出力、サンプルサイズ計算法が明確な形で提供されていないことである。

実際、最大対比法を利用している過去の例では、第 1 種の過誤確率や検出力の計算に正規近似あるいは単純なモンテカルロ・シミュレーションが用いられている。それは、これらの手法における p 値、検出力、サンプルサイズの計算には、非心多変量 t 分布の確率計算、より一般には特異な相関行列 (singular correlation matrix) を持つ非心多変量 t 分布での確率計算が必要であるのに、これを実現する計算プログラムが明確な形で存在していないからである。

これに関連した過去の研究では、岸本・浜田 (1994) が相関行列が正則な場合の最大対比法の p 値計算に関して、4 次元までの計算を行うプログラムを示しているが、これは正規近似を用いたものである。SAS の probmc 関数を用いれば、各群のサンプルサイズが任意の場合のダネット検定の p 値計算が可能であるが、チューキー検定とウィリアムズ検定 (Williams, 1971, 1972) については、各群のサンプルサイズが等しい場合しかその計算ができない。これに対して Genz and Bretz (1999) は、1 次独立な対比の組 (multiple contrasts) に対する確率計算の原理を提示して、そのためのプログラムをホームページ上に掲載している。これは任意の最大対比法における p 値、検出力、サンプルサイズ計算に適用することが可能なものであるが、これを実現するのは、データ解析の実際家にとってかなり困難な作業であると思われる。

そこで著者らは、Genz and Bretz (1999) の原理とサブルーチンを利用して、相関行列が特異な多変量 t 分布での確率計算が必要な、任意の最大対比法に適用できる p 値、検出力、サンプルサイズ計算プログラムを SAS/IML 上で作成した。以下にそのプログラムの原理と実行法を紹介する。すなわち第 2 章では本研究の対象とする最大対比法を定式化し、それに関する確率計算の原理を説明する。第 3 章では p 値、検出力、サンプルサイズ計算プログラムを説明する。第 4 章では現実的な状況を例にしてその利用法を紹介する。第 5 章では考察とまとめを行う。

2. 最大対比法

2.1 定式化

本論文では、水準に順序がある一元配置型データを対象とする。著者らは臨床試験や毒性試験に対する応用を意識しているので、水準が薬物の用量に対応しており、各水準での複数の観測値はヒトあるいは動物の個体に対応していると想定する。その関係で、各水準を「群」と言うことにする。群の数を K とする。第 i 群 ($i = 1, 2, \dots, K$) の各々に n_i 個の観測値 y_{ij} ; $j = 1, 2, \dots, n_i$ があり、 $\{y_{ij}; i = 1, 2, \dots, K, j = 1, 2, \dots, n_i\}$ は互いに独立に $N(\mu_i, \sigma^2)$ に従うものとする。平均 $\mu_1, \mu_2, \dots, \mu_K$ には、群の順序に対して単調増加という関係が想定できるものとし、その関係を「用量反応関係」と言うことにする。

以下の議論では、帰無仮説として用量反応関係がないこと、すなわち $H_0: \mu_1 = \mu_2 = \dots = \mu_K$ を想定し、それを臨床試験や毒性試験でよく現れる用量反応関係に対して検定することを考える。この帰無仮説 H_0 を永田・吉田 (1997) に倣って「包括的帰無仮説」(overall null hypothesis) と言うことにする。

この型のデータと $\sum_{i=1}^K c_i = 0$ となる定数 $c_1, c_2, \dots, c_K$ に対して、式 (1) で定義される統計量を「対比ベクトル $\mathbf{c} = (c_1, c_2, \dots, c_K)'$ に対する対比統計量」と言うことにする．上つきの「'」は転置である．

$$t = \frac{\mathbf{c}' \bar{\mathbf{y}}}{\sqrt{\hat{\sigma}^2 \mathbf{c}' \mathbf{M} \mathbf{c}}} \quad (1)$$

ただし、

$$\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_K)', \quad \bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, \quad (2)$$

$$\hat{\sigma}^2 = \frac{1}{n-K} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2, \quad n = \sum_{i=1}^K n_i, \quad (3)$$

$\mathbf{M}$ は第 i 対角要素が $1/n_i$ の対角行列である．

よく知られているように、この統計量は、帰無仮説 $H_0: \mathbf{c}'\boldsymbol{\mu} = 0$ を対立仮説 $H_1: \mathbf{c}'\boldsymbol{\mu} > 0$ に対して検定するための最良な検定統計量である．ただし、 $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_K)'$ である．帰無仮説 H_0 の下で t は自由度 $n-K$ の t 分布に従うので、有意水準 α の検定を行うには、 $t > t(n-K, \alpha)$ のときに H_0 を棄却すればよい．ただし、 $t(n-K, \alpha)$ は自由度 $n-K$ の t 分布の上側 $100\alpha\%$ 点である．

用量反応関係の関数形が、直線というように固定できる場合には、 r を正整数として、

$$\mathbf{c} = \begin{cases} (-r, -r+1, \dots, -1, 0, 1, \dots, r-1, r)' & K = 2r+1 \text{ のとき} \\ (-2r+1, -2r+3, \dots, -1, 1, \dots, 2r-3, 2r-1)' & K = 2r \text{ のとき} \end{cases}$$

というように対比ベクトルを定め、これに対する対比統計量を検定統計量にするのが合理的である．

しかし、臨床試験や毒性試験では、いろいろな型の用量反応関係に対して万遍なく高い検出力を保持できる検定法が望ましい．そこで、複数の用量反応関係のそれぞれに対応する対比ベクトル $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m$ を考え、これらに対する対比統計量 $t_1, t_2, \dots, t_m$ を用意し、その最大値

$$t_{\max} = \max_{l=1,2,\dots,m} t_l \quad (4)$$

を検定統計量にすることが考えられる．このようにして $t_{\max}$ がある値以上のときに帰無仮説を棄却する検定を、Yoshimura et al. (1997) に倣って「最大対比法」と言うことにする．

最大対比法では複数の対比ベクトルを用いる．これを $m \times K$ 行列の形式に並べた対比係数行列 $\mathbf{C}$ を次のように定める．

$$\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m)' = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1K} \\ c_{21} & c_{22} & \cdots & c_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mK} \end{pmatrix}. \quad (5)$$

たとえば $K = 5$ 群のときのダネット検定や $K = 4$ 群でのチューキー検定の対比係数行列 C_{DT} , C_{TK} は, それぞれ式 (6), (7) で与えられる.

$$C_{DT} = (c_1, c_2, \dots, c_4)' = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (6)$$

$$C_{TK} = (c_1, c_2, \dots, c_6)' = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}. \quad (7)$$

対比係数行列 C を一つ定めると, 最大対比法が一つ定まるので, 以下では最大対比法を特徴づける対比係数行列を「最大対比法の定義行列」と言うことにする.

2.2 p 値と検出力の計算原理

最大対比法では, 包括的帰無仮説が真のときに第 1 種の過誤確率が有意水準 α を超えないように定めるものとする. これは多重比較法でのファミリーワイズエラー (FWE) の制御とは少し違うが, 臨床試験データや毒性試験データから, 薬効や毒性の検証を行う際の判定法に対しては合理的な制約条件である.

同じ視点から, この包括的帰無仮説の下での検定統計量の確率分布において, 観測された検定統計量の実現値 t_{max}^* より大きな値を取る確率 $p = \Pr(t_{max} > t_{max}^* | H_0) = 1 - \Pr(t_{max} \leq t_{max}^*) = 1 - \Pr(t_1 \leq t_{max}^*, t_2 \leq t_{max}^*, \dots, t_m \leq t_{max}^*)$ を本論文では「 p 値」と言い, 考えている対立仮説 H_1 の分布において棄却限界値 a より大きな値を取る確率 $\Pr(t_{max} > a | H_1) = 1 - \Pr(t_{max} \leq a)$ を検出力と言うことにする. この検出力定義も, 複数の帰無仮説のそれぞれに対立仮説があるとする, 多重比較法の枠組みでの検出力定義からみれば不十分なところがあるが, 上記の p 値定義と同様な意義を持っているので, 本論文ではこの定義に基づいて議論を進める.

この定義における p 値と検出力を求めるには, 一般に対比統計量の同時確率分布を考え, その分布の下で標本空間の棄却域での確率を計算すればよい. 前章での独立, 正規, 等分散の仮定に基づけば, 対比統計量 $\{t_l; l = 1, 2, \dots, m\}$ の同時分布は式 (8) の右辺の統計量の同時分布と同等である.

$$t_l = \frac{z_l}{\sqrt{\chi_{n-K}^2 / (n-K)}}, \quad l = 1, 2, \dots, m \quad (8)$$

ただし, z_l は平均が $c_l' \mu / \sqrt{\sigma^2 c_l' M c_l}$ の確率変数であり, $(z_1, z_2, \dots, z_m)'$ の分散共分散行列 R は第 (a, b) 要素が

$$r_{ab} = c_a' M c_b / \sqrt{(c_a' M c_a)(c_b' M c_b)} \quad (9)$$

で表される行列である．また， χ^2_{n-K} は， $z_1, z_2, \dots, z_m$ と独立に自由度 $(n-K)$ の χ^2 分布に従う確率変数である．分散共分散行列 R は対角線要素が 1 なので相関行列と言うこともできる．

t_l の周辺分布は自由度 $(n-K)$ ，非心度

$$\delta_l = \frac{c_l' \mu}{\sqrt{\sigma^2 c_l' M c_l}} \quad (10)$$

の非心 t 分布である．そこで以下では，Johnson and Kots (1972, p.132) に倣って， $(t_1, t_2, \dots, t_m)'$ の同時分布を自由度 $(n-K)$ ，相関行列 R ，非心ベクトル $\delta = (\delta_1, \delta_2, \dots, \delta_m)'$ の多変量 (m 変量) 非心 t 分布と言うことにする．ただし，自明のときは「非心」を省略するし，包括的帰無仮説の場合のように非心ベクトルがゼロベクトルのときは意識的に「非心」を省略する．

相関行列 R は最大対比法の定義行列のランクが m より小さいとき特異 (非正則) になる．このような場合には，多変量 t 分布の確率密度関数を明示的に書くことができないので， p 値・検出力の計算は， $\text{rank}(C)$ 次元の (退化していない) 同時確率密度関数に対して， t_{max} がある値より大きい領域を調べ，その領域に対する積分を行なうことで求めることになる．

例として $K=3$ ， $m=3$ ， $n_1=n_2=n_3$ ，最大対比法の定義行列が式 (11) で与えられる場合について，計算原理を説明しよう．

$$C = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \quad (11)$$

この定義行列は $\text{rank}(C) = 2$ であり，第 1 行，第 2 行に対応する対比統計量 t_1, t_2 に対して，第 3 行に対応する対比統計量は $t_3 = t_2 - t_1$ になる．したがって最大対比法の棄却域は図 1 のようになる．この領域の確率を (t_1, t_3) の分布，すなわち相関行列が

$$R = \begin{pmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{pmatrix} \quad (12)$$

の 2 変量 t 分布から計算すればよい．非心ベクトルがゼロベクトルのときにこの確率が丁度 α になるようにすれば棄却限界値 a が求められる．その a を使って，しかるべき非心度のときの棄却域の確率を計算したものが検出力である．

p 値を求めるには，検定統計量の観測値 t_{max}^* に基づいてこの 2 次元空間に $\{(t_1, t_3); t_1 > t_{max}^*\} \cup \{(t_1, t_3); t_1 + t_3 > t_{max}^*\} \cup \{(t_1, t_3); t_3 > t_{max}^*\}$ の領域を作り，そこでの確率を包括的帰無仮説の下で計算すればよい．

ここには単純な例を挙げたが，一般の最大対比法の受容域 (棄却域の補集合) は，多次元空間における凸多角形領域の積分計算となる．これを具体的な定義行列について定めれば， p 値，棄却限界値，検出力が求まる．さらに検出したい対立仮説の用量反応関係を定めて，必要な検出力を与えるサンプルサイズを逐次近似で求めれば，サンプルサイズ計算ができる．これが計算原理である．

図 1. 最大対比法の棄却域

3. 最大対比法における確率・サンプルサイズ計算プログラム

3.1 SAS 上でのプログラム sub_mvt.sas

任意の自由度, 非心ベクトル, 相関行列, 積分範囲に対して m 変量非心 t 分布 ($m = 100$) の確率積分を計算するには Bretz のホームページ “www.bioinf.uni-hannover.de/~bretz/” で公開されている SAS/IML サブルーチンが利用できる. このサブルーチンは相関行列が特異であっても差し支えない.

著者らは, このサブルーチンを取り入れ, さらに多くのサブルーチンを作成し, 任意の最大対比法の p 値, 検出力, サンプルサイズ計算のための SAS プログラム「sub_mvt.sas」を作成した. このプログラムは非常に大きいので, 本論文では付録にも掲載していない. そこで, インターネット経由で取得できるように, 東京理科大学「医薬統計コース」ホームページ “www.rs.kagu.tus.ac.jp/yoshilab/iyaku/top.html” の「公開プログラム」の欄に圧縮ファイルで置いた. これを取得解凍し参照されているものとして, 5 群の場合のダネット検定を例にして, 次節において内容と使い方を説明する. その際に例示したプログラムは, MCM.sas という名前で, sub_mvt.sas と一緒に圧縮ファイル内に置いてある.

公開プログラム利用における慣例に従って, 本プログラムを利用したときは, プログラムの作成者が西山智であることを明記していただきたい.

3.2 p 値 計 算 例

5 群のダネット検定の定義行列は式 (6) の通りである. 5 群のサンプルサイズがすべて 10 であり, 検定統計量の実現値が 2.0 であったときの p 値計算プログラムと, それを用いたときの出力結果を表 1 に示す.

表 1. p 値計算例

<pre> PROC IML; %include '(sub_mvt.sas の保存先パス)¥sub_mvt.sas'; N={10 10 10 10 10}; t=2.0; Contrast={-1 1 0 0 0, -1 0 1 0 0, -1 0 0 1 0, -1 0 0 0 1}; ABSEPS=0.0001; RUN PVAL_MCM(ABSEPS,CONTRAST,t,N,PVAL); QUIT; </pre>			
出力結果:			
DIM	ERROR	PVAL	INFORM
4	0.00005	0.0789269	0

このプログラムで，“%include '(sub_mvt.sas の保存先パス)¥sub_mvt.sas';” は、著者らが作成したプログラムを呼び込む指示である．“N=” には，“N={ n_1 n_2 ... n_K };

” というように、各群のサンプルサイズ n_i ; $i = 1, \dots, K$ を列挙する．各値の間はスペースで区切らなければならない．“t=” のところには実現値 2.0 を入れる．

“Contrast=” には、最大対比法の定義行列を、各行を「 」で区切り、列間にスペースをおいて入れる．今回は 5 群なので式 (6) の定義行列を入れる．“ABSEPS=” には、求めたい計算精度を、計算値の標準誤差の 3 倍の値で、“ABSEPS=0.0001;” のような形で入れる．精度を標準誤差で指定するのは、このプログラムが準乱数を用いて関数値の計算点を選び数値積分をしているので、計算値にランダム性が含まれるためである．“PVAL_MCM” は使用しているプログラム「sub_mvt.sas」内のサブルーチン名である．

このプログラムは、Bretz and Genz のサブルーチンの制約によって、式 (9) で与えられる相関行列が半正定値行列 (positive semi-definite matrix) でなければならない．

出力結果において、“DIM” には対比の個数 m ，“ERROR” には推定値の 99% 信頼区間幅，“PVAL” には計算した p 値，“INFORM” には計算が求められた精度で実行されたかどうかが表示される．出力が 0 であれば問題ないが、1 であれば“ABSEPS” の計算精度が達成できていないことを意味する．もし、制約を犯していたりその他の異常が生じていれば 0,1 以外の値が出力される． m が 100 を超えている場合は 2，相関行列が半正定値行列でない場合は 4 である．

3.3 検出力計算例

前節と同じ 5 群のダネット検定で有意水準が 2.5%，各群のサンプルサイズが (30, 15, 15, 15, 15)，対立仮説の平均ベクトルが (0.0, 0.25, 0.50, 0.75, 1.0)， $\sigma^2 = 1.0$ という場合について、検出力計算プログラムとその出力結果を表 2 に示す．

表 2. 検出力計算例

```

PROC IML;
  %include '(sub_mvt.sas の保存先パス)¥sub_mvt.sas';
  Alpha=0.025;
  Contrast={-1 1 0 0 0,
            -1 0 1 0 0,
            -1 0 0 1 0,
            -1 0 0 0 1};
  Expect={0 0.25 0.50 0.75 1.00};
  VARIANCE=1;
  N1=30; N_ALLOC={2 1 1 1 1};
  Switch=1;
  Eps1=0.0001; ABSEPS=0.0001;
  RUN ESTPOWER(Switch,Alpha,N1,Eps1,ABSEPS,Expect,
               VARIANCE,CONTRAST,N_ALLOC,N,Power,CriVal);
  IF Switch=1 THEN PRINT N CriVal Power ; ELSE PRINT N CriVal ;
QUIT;

```

出力結果:

N					CRIVAL	POWER
30	15	15	15	15	2.5195313	0.8209445

すでに説明したことを省くと各項目の意味・使い方は次の通りである。

プログラムにおいて, “Alpha=” には設定する有意水準を入れる. “Expect=” と “VARIANCE=” には, 検出力を計算する対立仮説の平均ベクトルと分散を入れる. “N1=” には第 1 群のサンプルサイズを入れ, “N_ALLOC=” には第 1 群に対する各群のサンプルサイズの比を “N_ALLOC={2 1 1 1 1};” というように入れる. 比は整数でなくても良い.

p 値計算のときは検定の棄却限界値が不要であるが, 有意水準を与えて検出力を計算するときは, 棄却限界値の計算が必要である. これは逐次近似で求めるので, “EPS1=” には, 与えた有意水準と逐次近似で計算した第 1 種の過誤確率との差として, 求めたい精度の値を入れる. このプログラムではその精度を 0.0001 に設定してある.

“Switch=” には, 棄却限界値と検出力の両方を出力するとき “Switch=1;” を指定し, 棄却限界値のみを出力するときは, 1 以外の値を入れる.

“ESTPOWER” はプログラム内のサブルーチン名である.

上記のプログラムを実行すると出力結果の “N” には, 各群のサンプルサイズが, 30, 15, 15, 15, 15 というように出力され, “CRIVAL” と “POWER” には計算結果の棄却限界値 2.5195313 と検出力 82.09445% が出力される.

表 3. 必要サンプルサイズ計算例

```

PROC IML;
  %include '(sub_mvt.sas の保存先パス)¥sub_mvt.sas';
  Alpha=0.025; Beta=0.2;
  Contrast={-1 1 0 0 0,
            -1 0 1 0 0,
            -1 0 0 1 0,
            -1 0 0 0 1} ;
  Expect={0 0.25 0.50 0.75 1.00};
  VARIANCE=1;
  N_ALLOC={2 1 1 1 1};
  Switch=1;
  Eps1=0.0001; ABSEPS=0.0001;
  Nmin=5; Nmax=100; Eps2=0.01;
  DIFF=1; N1=Nmax;
  DO UNTIL (ABS(Power-(1-Beta))<Eps2);
    IF DIFF>0 THEN Nmax=N1; ELSE Nmin=N1;
    N1=(Nmax+Nmin)/2;
    RUN ESTPOWER(Switch,Alpha,N1,Eps1,ABSEPS,Expect,
                 VARIANCE,CONTRAST,N_ALLOC,N,Power,CriVal);
    DIFF=Power-(1-beta);
  END;
  PRINT N CriVal Power;
QUIT;

```

出力結果:

N	CRIVAL	POWER
28.75 14.375 14.375 14.375 14.375	2.5234375	0.8015946

3.4 サンプルサイズ計算例

各種の試験・実験に最大対比法を応用するときには、与えられた有意水準と検出力を保証するサンプルサイズの計算が重要である。前節と同じ5群のダネット検定で、有意水準2.5%、前節で示した対立仮説に対して検出力が80%になるサンプルサイズ計算プログラムとその出力例を表3に示す。

すでに説明したことを省くと各項目の意味・使い方は次の通りである。

“Beta=”には検出力に対応する第2種の過誤確率すなわち $1 - 0.80 = 0.20$ を入れる。このプログラムでは、“N_ALLOC=”に各群のサンプルサイズの比を入れ、第1群のサンプルサイズを計算するようにしてある。比は多くの場合1にするが、ここでは比が違っていても計算できることを示すために、1以外の比を指定している。すべての群のサンプルサイズを等しくするときは“N_ALLOC={1 1 1 1 1};”とすれば良い。

“Nmim=” と “Nmax=” に入れるのは、逐次近似における第 1 群のサンプルサイズの初期値で、それぞれ十分小さい値と大きい値の二つを指定する。“Eps2=” には、逐次近似で求める精度について、指定した検出力と計算した検出力の差の限界値を与える。ここでは 0.01 を入れている。

“DO UNTIL ~ END” は、初期値の下限と上限の間で必要なサンプルサイズを求めるためのプログラム、その前の “DIFF=1”, “N1=Nmax” は、そのプログラムのための初期値であり、然るべき条件を満たせば他の値でも差し支えないが、このままにしておく方が無難である。

サンプルサイズと検出力の関係は単調であるから、この程度の簡単なアルゴリズムで十分早い収束が達成できる。この例では 2 回の反復、計算時間 5.87 秒で計算が収束している。

出力結果では、サンプルサイズが小数で出力されるが、実際はその小数を下回らない整数にしなければならない。この例では、各群のサンプルサイズを 30, 15, 15, 15, 15 にすることになる。このときの検出力が前節で例示した 82.1% で、出力として示された小数のサンプルサイズに対する検出力計算値 80.2% より大きくなる。

4. 最大対比法の定義行列を比較するためのプログラム利用例

4.1 比較する定義行列

各種の試験・実験に最大対比法を適用するには、最大対比法の定義行列をどのように選択するかが問題となる。本章では、第 II 相臨床試験に採用され得るデザインを想定して、最大対比法で有効性を評価するとしたときの定義行列の比較を考える。

状況として、第 1 群がプラセボ群 (C) すなわち用量が 0 の群、第 2 群が低用量群 (L)、第 3 群が高用量群 (H) の 3 群並行比較臨床試験を考える。この場合、用量反応関係が単調増加であることは前提にできるが、用量反応関係がどのような関数形であるかは確かでないのが普通である。そこで被験薬がプラセボに比べて有効であることを検証するために、最大対比法を使うことが考えられる。しかしどのような最大対比法、言いかえればどのような定義行列を採用するのが良いかは自明ではない。検出力を比較して検討すべきである。

例えば Stewart and Ruberg (2000) は、用量反応形状が未知である場合には Half-up (HU) と Half-down (HD) の二つの対比を用いた最大対比法 (HU/HD 法) が必要サンプルサイズに関してミニマックス的に良い検定法であると結論している。しかしその検討は、5 群の場合に限定されていて、しかも多変量 t 分布ではなくその正規近似を用いて議論しているので、その結論を一般的なものとして受け入れるわけにいかない。ここで、 K 群のときの対比 HU と対比 HD の対比ベクトルは、それぞれ、

$$\begin{aligned} c_{HU} &= K(0, K-1, K-1, \dots, K-1)' + K(0, 1, 2, \dots, K-1)' - 0.5(K-1)(3K-2)(1, 1, 1, \dots, 1)', \\ c_{HD} &= K(0, 0, 0, \dots, K-1)' + K(0, 1, 2, \dots, K-1)' - 0.5(K-1)(K+2)(1, 1, 1, \dots, 1)' \end{aligned}$$

で与えられるものである。前者は低用量で反応が急増し、その後、高用量に向かってなだらかに増加するような (急上昇型) 用量反応曲線を良く検出する対比であり、後者はその逆の (緩上昇型) 用量反応曲線を良く検出する対比である。

HU と HD をさらに極端にすると、低用量で反応が十分大きくなって高用量では低用量で変わらないという Upper-bound 型 (UB, 天井型) と、その逆の Lower-bound 型 (LB, 底型) の用

量反応曲線を良く検出する対比が考えられる．もちろん，直線的な用量反応曲線（LR，直線型）を良く検出する対比も考えられる．

これらを組み合わせると，たとえば以下に示す 5 通りの定義行列が考えられる．これらについて，その検出力や必要サンプルサイズがどのようなときにどのように違うかを検討しよう．便宜上それぞれの定義行列を最大対比法の名前にして，LR 法，DT 法などと呼ぶことにする．

$$C_{LR/LB} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & -1 & 2 \end{pmatrix}, \quad (13)$$

$$C_{LR} = \begin{pmatrix} -1 & 0 & 1 \end{pmatrix}, \quad (14)$$

$$C_{HU/HD} = \begin{pmatrix} -7 & 2 & 5 \\ -5 & -2 & 7 \end{pmatrix}, \quad (15)$$

$$C_{DT} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad (16)$$

$$C_{LR/UB} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 1 & 1 \end{pmatrix}. \quad (17)$$

4.2 数値計算による比較

3 群の場合について，用量反応曲線としてあり得る 5 つの形状（図 2）を考え，前節の定義行列で定められる 5 つの最大対比法で検出力 90% を保証するサンプルサイズを計算すると，表 4 が得られる．いずれも，有意水準は 2.5% で，用量反応曲線の上下方向の最大差が $\Delta = 1$ ，誤差分散が $\sigma^2 = 5$ の場合である．

LR 法の必要サンプルサイズは， Δ だけで決まるのでどの用量反応曲線に対しても同じである．DT 法（ダネット検定）は天井型を除くとサンプルサイズが大きくなる．これは用量反応曲線が単

図 2. 例題において検出力を求める用量反応関係

表 4. 1 群あたり必要サンプルサイズ (実際の検出力)

用量反応関係	最大対比法				
	LR/LB	LR	HU/HD	DT	LR/UB
天井型	117 (0.9008)	106 (0.9008)	93 (0.9006)	90 (0.9001)	85 (0.9032)
急上昇型	116 (0.9000)	106 (0.9008)	104 (0.9000)	110 (0.9000)	101 (0.9018)
直線型	112 (0.9006)	106 (0.9008)	111 (0.9045)	121 (0.9020)	112 (0.9003)
緩上昇型	101 (0.9022)	106 (0.9008)	104 (0.9001)	123 (0.9009)	116 (0.9000)
底型	85 (0.9032)	106 (0.9008)	93 (0.9010)	123 (0.9001)	117 (0.9008)

調増加という特徴を使っていないからである．LR/UB 法は天井・急上昇型（上に凸）に有利で，逆に LR/LB 法は緩上昇・底型（下に凸）に有利である．HU/HD 法は，Stewart and Rubwerg (2000) が主張するように，かなりロバストである．著者らの経験では，この種の臨床試験には上に凸型が多いので，この計算結果から，LR/UB 法を使うのが良いと結論するのが合理的であろう．

5. 考察とまとめ

5.1 多変量 t 分布について

K 群の一元配置での 1 次独立な対比の数は，最大で $K-1$ である．最大対比法を構成する対比統計量がその数以下で，しかも対応する対比ベクトルが 1 次独立であれば， p 値，検出力計算は多変量 t 分布における矩形領域積分で計算できる．本論文で問題としている最大対比法では，しばしばこの 1 次独立性が失われ，複数の対比統計量の相関行列が特異になる．このときに対比統計量の組の確率分布を，正規分布の例に倣って「退化多変量 t 分布」(degenerated multivariate t -distribution) と呼ぶべきかどうかについては，著者らに迷いがあった．そこで文献検索を行ってみたところ，そういう表現は見あたらなかった．

実際のプログラムでの積分は，正則な相関行列をもつ多変量 t 分布に基づいて確率密度関数を考え，積分領域を凸多角形として行っているから，あえて退化多変量 t 分布という名前は入れなかったが，この積分をオブジェクト化する場合は退化多変量 t 分布という概念で整理した方が分かりやすいかもしれない．これについては関心を持つ研究者の間での合意が必要であろう．

5.2 最大対比法と多重比較法

最大対比法を構成する複数の対比統計量は，それぞれ対応する帰無仮説に対する検定統計量になっている．従って最大対比法は，複数の帰無仮説を持った多重比較法と見ることもできる．もしそう見ると，多重比較法の性能の議論でよく知られているように，第 1 種の過誤確率をファミリーワイズエラー (family wise error; FWE) として制御することが考えられる．ところが本論文ではそのような見方をあえて採用せず，包括的帰無仮説のみに対して p 値を定義し，その計算を考えた．それは本論文が応用対象としている試験・実験データに関しては，用量反応関係が全くないかどうかに関心があって，各対比に対応する帰無仮説の個別検定には関心がないと考えたからである．もし，個別の帰無仮説を考えるなら，多重比較法の枠組みでの p 値定義をしなければならなくなるが，これは本論文の対象外である．

検出力計算とサンプルサイズ設計を行なう場合は，どのような対立仮説を想定するかが問題に

なるが、これについても現実的な応用を考え、個々の帰無仮説に対する非心度を設定するのではなく、ある用量反応関係を設定し、これに対する検出力を計算するようにした。多重比較法の枠組みで、多様な検出力定義に基づいて議論を行うことも、現実の応用においては重要である。しかしこの課題は、本論文とは別の機会に行うつもりである。

5.3 例示プログラムの性能

表 4 の計算には、原則として本論文のプログラムを用いているが、LR 法だけは SAS に用意されている関数 (probt, tinv) を利用している。単純な 1 変量 t 分布の計算だからである。

本論文で提案したプログラムの計算の速さと計算精度を確かめるために、表 4 と同じ内容を、100 万回反復の単純なモンテカルロ・シミュレーションと比較してみた。表 5 に示すのはそのモンテカルロ・シミュレーションで求めた第 1 種の過誤確率と検出力である。計算精度は提案プログラムと同程度であるから、このモンテカルロ・シミュレーションと本プログラムでの計算速度を比べてみよう。

計算時間は本プログラムの場合、表 4 の 1 つの検出力を求めるのに 1.0 ~ 6.5 秒であるのに対し、モンテカルロ・シミュレーションではそれが 281 ~ 388 秒であった。計算機が「Intel® Pentium®4 CPU 2.80 GHz メモリ 1.5 GB」の PC の場合である。

シミュレーション回数を 10 万回に減らした場合でも 30 ~ 40 秒かかるので、提案プログラムは単純なモンテカルロ・シミュレーションよりはるかに有効であることが分かる。

表 5. 100 万回の単純モンテカルロ・シミュレーションによる第 1 種の過誤・検出力

用量反応関係	最大対比法									
	LR/LB		LR		HU/HD		DT		LR/UB	
	第 1 種 の過誤	検出力	第 1 種 の過誤	検出力	第 1 種 の過誤	検出力	第 1 種 の過誤	検出力	第 1 種 の過誤	検出力
天井型	0.0249	0.9007	0.0250	0.9003	0.0251	0.9002	0.0251	0.9004	0.0251	0.9026
急上昇型	0.0251	0.9000	0.0250	0.9005	0.0249	0.8998	0.0248	0.8999	0.0251	0.9020
直線型	0.0252	0.9002	0.0250	0.9006	0.0248	0.9045	0.0249	0.9022	0.0250	0.9000
緩上昇型	0.0251	0.9021	0.0251	0.9002	0.0250	0.8999	0.0249	0.9004	0.0250	0.8999
底型	0.0248	0.9031	0.0249	0.9009	0.0254	0.9008	0.0250	0.8998	0.0250	0.9007

なお、表 1 のダネット検定の p 値を SAS の probmc 関数で計算した結果は 0.0789354 であった。作成したプログラムの計算結果と小数点 4 桁まで一致している。これも提案プログラムが十分な精度を持っていることの一つの証拠と言えよう。

謝 辞

本論文をまとめるに当たっては、2 人の査読者及び編集委員長から有益な示唆・助言を頂いた。ここに記して感謝の意を表する。

参考文献

- Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, **50**, 1096-1121.
- Jpn J Biomet Vol.24, No.2, 2003

- Genz, A. and Bretz, F. (1999). Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation*, **63**, 361-378.
- Hochberg, Y. and Tamehane, A. C. (1987). *Multiple Comparison Procedures*. New York: John Wiley & Sons, Inc.
- Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics : Continuous Multivariate Distributions*. New York: John Wiley & Sons, Inc.
- 岸本淳司, 浜田知久馬 (1994). 任意の対比群について多重比較を行う数値積分プログラム. 計算機統計学, **7**, 147-154.
- Kramer, C. Y. (1956). Extension of multiple range test to group means with unequal numbers of replications. *Biometrics*, **12**, 307-310.
- 永田靖, 吉田道弘 (1997). 統計的多重比較法の基礎. サイエンティスト社.
- Stewart, W. H. and Ruberg, S. J. (2000). Detecting dose response with contrasts. *Statistics in Medicine*, **19**, 913-921.
- Tukey, J. W. (1953). *The problem of multiple comparisons*. Mimeographed monograph.
- Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*, **27**, 103-117.
- Williams, D. A. (1972). The comparison of several dose levels with a dose zero control. *Biometrics*, **28**, 519-531.
- Yoshimura, I., Wakana, A. and Hamada, C. (1997). A performance comparison of maximum contrast methods to detect dose dependency. *Drug Information Journal*, **31**, 423-432.