氏名（本籍）	川崎玉恵（東京都）
学位の種類	博士（理学）
学位記番号	甲第1103号
学位授与の日付	平成28年3月18日
学位授与の要件	学位規則第4条第1項該当
学位論文題目	**Tests for mean vectors with two-step monotone missing data or unequal covariance matrices** **（2-ステップ単調欠測データまたは異なる共分散行列をもつ平均ベクトルの検定）**

論文審査委員　（主査）教授　瀬尾　隆
教授　宮岡　悦良　　教授　矢部　博
准教授　小谷　佳子　　准教授　橋口　博樹

論文内容の要旨

本論文では，2-ステップ単調欠測データをもつ平均ベクトルの検定統計量に対するバイアス修正変換統計量，及び異なる分散共分散行列をもつ平均ベクトルの検定統計量に対する自由度調整法を用いた近似解について考える．

第1章は序論であり，各テーマに対する研究の背景や方針，本論文の構成について述べている．多変量統計解析において，平均ベクトルの検定に関する研究は基礎的な問題である．本論文では2つのテーマに焦点を当てて議論を行う．1つめは，1標本問題における平均ベクトルの検定を欠測データのもとで議論する．統計的手法や理論で扱うデータは，欠測値を含まないことを仮定して議論されている場合が多いが，現実問題で扱うデータの多くは欠測値を含むデータである．本論文では，欠測構造のひとつである2-ステップ単調欠測データを仮定する．2-ステップ単調欠測データに対する最尤推定量を確率展開することで検定統計量の展開結果を与え，その1次，2次モーメントをそれぞれ導出する．さらにこれらの結果を用いて，2種類のバイアス修正変換統計量を与えている．2つめは，分散共分散行列が異なる2つの母集団における，平均ベクトルの検定について議論する．またデータには欠測値のない，完全データを仮定している．2標本問題において母集団の分散共分散行列が異なるとき，検定統計量の正確な分布の導出は難しいことが知られている．本論文ではF分布の自由度を調整することにより，検定統計量に対する近似分布を与える．さらに，バイアスを修正した近似分布も与えている．

第2章では，1つめのテーマである1標本問題における平均ベクトルの検定に対し，2-

ステップ単調欠測データのもとでのバイアス修正変換統計量について議論する．n_1 個の $p(=p_1+p_2)$ 変量標本ベクトル $\boldsymbol{x}_j^{[1]}=(\boldsymbol{x}_{1j}^{[1]\prime},\boldsymbol{x}_{2j}^{[1]\prime})'$, $j=1,2,\ldots,n_1$ が $N_p(\boldsymbol{\mu},\Sigma)$ に従い，$n_2(=n-n_1)$ 個の p_1 変量標本ベクトル $\boldsymbol{x}_j^{[2]}$, $j=1,2,\ldots,n_2$ が $N_{p_1}(\boldsymbol{\mu}_1,\Sigma_{11})$ に従っているとする．ただし

$$\boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix},\quad \Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$$

と分割する．このようなデータセットのことを2-ステップ単調欠測データと呼ぶ (Kanda and Fujikoshi (1998) 等を参照)．１標本問題における平均ベクトルの検定 $H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0$ vs. $H_1:\boldsymbol{\mu}\neq\boldsymbol{\mu}_0$ に対する検定統計量として，次のようなホテリングの T^2 型統計量を用いる．

$$T^2=(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)'\begin{pmatrix}(1/n)\widehat{\Sigma}_{11} & (1/n)\widehat{\Sigma}_{12}\\ (1/n)\widehat{\Sigma}_{21} & \widehat{\mathrm{Cov}}[\widehat{\boldsymbol{\mu}}_2]\end{pmatrix}^{-1}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0)$$

ただし，$\boldsymbol{\mu}_0$ は既知ベクトルとし，$\boldsymbol{\mu},\Sigma$ の最尤推定量を $\widehat{\boldsymbol{\mu}}$, $\widehat{\Sigma}$ とする．このときの最尤推定量は Anderson and Olkin (1985) によって与えられており，T^2 型統計量に対する近似上側パーセント点は，Chang and Richards (2009) や Seko et al. (2012) などで議論されている．本論文では，$n,n_1\to\infty$ かつ，$N_i\to\infty$ のとき $\gamma_i=N_i/N$ が正の定数に収束するもとで，$\widehat{\boldsymbol{\mu}}$ と $\widehat{\Sigma}$ の展開を摂動法によって与え，その結果を用いて T^2 型統計量，T^2 型統計量の２乗の統計量，T^4 型統計量の展開を与えた．ただし，$N=N_1+N_2$, $N_i=n_i-1$, $i=1,2$ とする．さらに，T^2,T^4 型統計量の展開結果の期待値を，次の定理によって与えた．

定理 2.1. T^2,T^4 型統計量の期待値はそれぞれ以下のように与えられる．

$$\mathrm{E}[T^2]=p+\frac{1}{N}c_1+\mathrm{O}(N^{-2}),\quad \mathrm{E}[T^4]=p(p+2)+\frac{1}{N}c_2+\mathrm{O}(N^{-2})$$

ただし

$$N=N_1+N_2,\quad c_1=p_1(p+4)+\frac{p_2(p+3)}{\gamma_1},$$
$$c_2=2p_1\{4(p_1+2)+(p_1+1)(p+2)+p_2(p+7)\}+\frac{2p_2}{\gamma_1}\{p_2+2+(p+2)(p+3)\}.$$

以上の結果を用いて，バイアス修正変換統計量を２つ提案する．１つめは，バートレット変換統計量を2-ステップ単調欠測データに適応させたもので，以下のように与えられる．

$$T_B^2=\left(1-\frac{1}{Np}c_1\right)T^2$$

ただし，c_1 は定理2.1における係数である．ここでバートレット変換とは，変換後の統計量の平均が極限分布であるカイ２乗分布の平均と $\mathrm{O}(N^{-1})$ まで一致するような変換である．T_B^2 に対しても，次の結果が得られた．

$$\mathrm{E}[T_B^2]=p+\mathrm{O}(N^{-2}),\quad \mathrm{E}[T_B^4]=p(p+2)+\frac{1}{N}\{c_2-2c_1(p+2)\}+\mathrm{O}(N^{-2})$$

２つめは，Fujikoshi (2000) による修正バートレット変換統計量を2-ステップ単調欠測データのもとで考える．Fujikoshi (2000) では完全データのもとで，変換後の統計量の１次，２次モーメントが，極限分布であるカイ２乗分布の１次，２次モーメントとそれぞれ $\mathrm{O}(N^{-1})$ まで一致するような変換を与えている．このような変換統計量を2-ステップ単調欠測データのもとで，以下のように与えた．

$$T_{MB}^2 = (N\beta_1 + \beta_2)\log\left(1 + \frac{T^2}{N\beta_1}\right)$$

ただし，β_1 と β_2 はそれぞれ

$$\beta_1 = \frac{2p(p+2)}{c_2 - 2c_1(p+2)},\ \beta_2 = \frac{(p+2)\{c_2 - 2c_1(p+4)\}}{2\{c_2 - 2c_1(p+2)\}}$$

であり，c_1, c_2 はそれぞれ，定理 2.1 における係数である．また，T_{BC}^2 に対しても，次のような結果が得られた．

$$\mathrm{E}[T_{MB}^2] = p + \mathrm{O}(N^{-2}),\ \ \mathrm{E}[T_{MB}^4] = p(p+2) + \mathrm{O}(N^{-2}) = 2p + \mathrm{O}(N^{-2})$$

ただし，$\beta_1 > 0$ かつ $N\beta_1 + \beta_2 > 0$ である．シミュレーションでは，１次，２次モーメント，分散，MSE について，漸近展開の結果と共に T^2, T_B^2, T_{MB}^2 をモンテカルロ・シミュレーションによって評価した．シミュレーションによる T^2 の値と漸近展開による近似の結果は，値が安定的に近いことが示され，T_B^2 と T_{MB}^2 はカイ２乗分布への収束を早めた変換であるため，シミュレーションでもその修正された近似精度の良さを示している．さらに，次元数に対して標本数が小さい場合にも同様の傾向が見られる．また，シミュレーションによる T^2 と T_B^2, T_{MB}^2 の上側 100α パーセント点を与え，帰無仮説における極限分布である自由度 p のカイ２乗分布で棄却するときの type I error を計算し，近似精度を議論した．検定統計量に対して理論的にモーメントの収束を早める変換を行った T_B^2 や T_{MB}^2 は，実験を行ったすべての場合において，高精度の近似値であることを示している．また，T_B^2 の近似値の精度も良いが，２次モーメントの収束も早めた T_{MB}^2 の方が，より精度が良いことが示されている．

第３章では，２標本問題における平均ベクトルの検定を母集団の分散共分散行列が等しくない場合で議論する．母集団の分散共分散行列が等しくない場合の平均ベクトルの検定は一般に多変量 Behrens-Fisher 問題と呼ばれており，多くの議論がなされている．たとえば James (1954) は，検定統計量の Cornish-Fisher 展開によるカイ２乗分布を用いた近似解を与え，また Yao (1965) は検定統計量の分布について，自由度を調整することによって近似が良くなるような F 分布を与えている．また，Yanagihara and Yuan (2005) では F 近似による自由度を調整した近似解，修正バートレット補正におけるカイ２乗近似解を与えている．本論文では，Yanagihara and Yuan (2005) における近似法を
より精密にした結果を与えた．$n^{(i)}$ 個の標本ベクトル $\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(i)}, \ldots, \boldsymbol{x}_{n^{(i)}}^{(i)}$ がそれぞれ独立に $N_p(\boldsymbol{\mu}^{(i)}, \Sigma^{(i)})$，$i = 1, 2$ に従うとする．このとき，２標本問題における平均ベクトルの検定 $H_0 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$ vs. $H_1 : \boldsymbol{\mu}^{(1)} \neq \boldsymbol{\mu}^{(2)}$ を，母集団の分散共分散行列が等しくない $\Sigma^{(1)} \neq \Sigma^{(2)}$ の場合で考える．ここで，母集団の分散共分散行列が等しい $\Sigma^{(1)} = \Sigma^{(2)}$ の場合の平均ベクトルの検定は，検定統計量としてホテリングの T^2 統計量がよく知られており，その分布は帰無仮説のもとで正確に F 分布に従う．母集団の分散共分散行列が等しくない場合の帰無仮説のもとでの正確な分布は，Nel et al. (1990) や Girón and Castillo (2010) などによって与えられているが，これらは計算的に扱いにくいものとなっている．

本論文ではこの仮説検定問題において，Welch (1938) によって与えられた１変量の統計量を多変量に拡張した，次の検定統計量を用いる．

$$T = (\overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)})'\left(\frac{S^{(1)}}{n^{(1)}} + \frac{S^{(2)}}{n^{(2)}}\right)^{-1}(\overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)})$$

ただし，$\bar{\boldsymbol{x}}^{(i)}$ を第 i 母集団の標本平均ベクトル，$S^{(i)}$ を第 i 母集団の不偏標本分散共分散行列とする．ここで，この T 統計量は

$$n = n^{(1)} + n^{(2)},\ \ \boldsymbol{z} = \sqrt{\frac{n^{(1)}n^{(2)}}{n}}\overline{\Sigma}^{-\frac{1}{2}}(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)}),\ \overline{\Sigma} = \frac{n^{(2)}}{n}\Sigma^{(1)} + \frac{n^{(1)}}{n}\Sigma^{(2)},$$
$$W = \overline{\Sigma}^{-\frac{1}{2}}\left(\frac{n^{(2)}}{n}S^{(1)} + \frac{n^{(1)}}{n}S^{(2)}\right)\overline{\Sigma}^{-\frac{1}{2}},\ U = \frac{\boldsymbol{z}'\boldsymbol{z}}{\boldsymbol{z}'W^{-1}\boldsymbol{z}}$$

を用いて

$$T = \boldsymbol{z}'W^{-1}\boldsymbol{z} = \frac{\boldsymbol{z}'\boldsymbol{z}}{U}$$

と書き表すことができ，さらに U の近似分布を未知パラメータ ν, ϕ を用いて χ^2_ν/ϕ と考えると，$\mathrm{E}[U]$, $\mathrm{E}[U^2]$ の近似値はそれぞれ ν/p, $\nu(\nu+2)/\phi^2$ と表すことができる．よって $\boldsymbol{z}'\boldsymbol{z}$ は自由度 p のカイ２乗分布に正確に従うので，T 統計量を近似的に自由度が異なるカイ２乗分布の比で表せると考えれば，$\nu T/(p\phi)$ の近似分布は自由度 p, ν の F 分布とみなすことができる．そこで，U における W^{-1} を摂動展開することで $\mathrm{E}[U]$ と $\mathrm{E}[U^2]$ の展開を N^{-2} の項まで求めることができ，展開結果を次の定理によってそれぞれ与えた．ただし，$N = n-2$ とする．

定理 3.1. $\mathrm{E}[U]$ の N^{-2} までの展開結果は以下のように与えられる．

$$\mathrm{E}[U] = 1 - \frac{\theta_1}{N} + \frac{1}{N^2}(\theta_2 - \theta_3) + O(N^{-3})$$

定理 3.2. $\mathrm{E}[U^2]$ の N^{-2} までの展開結果は以下のように与えられる．

$$\mathrm{E}[U^2] = 1 - \frac{2}{N}(\theta_1 - \theta_4) + \frac{1}{N^2}(2\theta_5 - \theta_6) + O(N^{-3})$$

ただし，$\theta_1, \theta_2, \ldots, \theta_6$ は $\Sigma^{(i)}$ と $\overline{\Sigma}$ の関数である．よって，$\mathrm{E}[U]$, $\mathrm{E}[U^2]$ はそれぞれ ν/ϕ, $(\nu(\nu+2))/\phi^2$ と近似的に等しいと見なすことができるため，定理 3.1 と 3.2 の展開結果とを連立方程式で解くことにより，次のような ν と ϕ の新たな近似解を得ることができた．

$$\nu_S = \frac{2(N^2 - N\theta_1 + \theta_2 - \theta_3)^2}{N^2(N^2 - 2N\theta_1 + 2N\theta_4 + 2\theta_5 - \theta_6) - (N^2 - N\theta_1 + \theta_2 - \theta_3)^2},$$
$$\phi_S = \frac{N^2\nu_S}{N^2 - N\theta_1 + \theta_2 - \theta_3}$$

さらに，$\Sigma^{(i)}$ と $\overline{\Sigma}$ は未知であるため，$\theta_1, \theta_2, \ldots, \theta_6$ を計算するには推定量を用いる必要がある．そこで推定量を用いたときに生じるバイアスを修正した，次のような ν と ϕ も与えた．ただし，θ_1^* と θ_4^* は $\Sigma^{(i)}$ と $\overline{\Sigma}$ の関数であり，$\vartheta = N^2 - N\theta_1 + \theta_2 - \theta_3 - \theta_1^*$ とする．

$$\nu_{BC} = \frac{2\,\vartheta^2}{N^2\,(N^2 - 2N\theta_1 + 2N\theta_4 + 2\theta_5 - \theta_6 - 2\theta_1^* + 2\theta_4^*) - \vartheta^2},\ \phi_{BC} = \frac{N^2\nu_{BC}}{\vartheta}$$

また，本論文で与えた近似解 ν_S と ϕ_S，ν_{BC} と ϕ_{BC} を用いた F 分布による近似上側 100α パーセント点の精度と先行研究との比較をするため，モンテカルロ・シミュレーションによって type I error による近似精度を議論した．シミュレーションは，第２母集団の分散共分散行列 $\Sigma^{(2)}$ を p 次元の単位行列で固定し，第１母集団の分散共分散行列 $\Sigma^{(1)}$ に対して２つの構造を仮定する．１つめの構造は $\Sigma^{(1)} = \mathrm{diag}(\epsilon, \epsilon^2, \ldots, \epsilon^p)$, $\epsilon = 1, 5, 10, 20$ で，２つめの構造は $\Sigma^{(1)} = \sigma^2 I_p$, $\sigma^2 = 0.1(0.2)0.7, 2, 5, 10, 20, 30$ である．先行研究の Yanagihara and Yuan (2005) は，$\Sigma^{(1)}$ と $\Sigma^{(2)}$ が大きく乖離している場合や標本数が少ない場合では T 統計量のパーセント点の近似があまり良くないが，本結果を使うことでこれらの場合でも近似精度が改良されていることを示した結果となっている.

論文審査の結果の要旨

本論文は，１標本問題および２標本問題における平均ベクトルの検定問題について述べたものであり，３つの章から構成される.

第１章は序論である．まず，本論文の研究の位置づけと研究の背景について述べている．多変量統計解析の検定問題については多くの問題があるが，その中で平均ベクトルに関する検定問題は重要かつ基礎となる問題である．このような背景の下で，データに欠損が生じた場合の平均ベクトルの検定問題と分散共分散行列が異なる２つの平均ベクトルの同等性検定に注目し，本論文の研究テーマと目的を与えている.

第２章では，データが単調欠測という構造を持つ場合の１標本問題における平均ベクトルの検定問題を取り扱っている．まず，欠測データの構造を２ステップ単調欠測データに絞り，記号や仮説検定に使われる T^2 型検定統計量の定義および平均ベクトルと分散共分散行列の最尤推定量について説明している．そして，これらの結果を基に，摂動展開法を用いて，T^2 型検定統計量の１次モーメントと２次モーメントの漸近展開を N^{-1} の項まで漸近展開することに成功している．ただし，N は欠測データを含む全サンプル数である．特筆すべき点は，データが欠測している設定の下で摂動展開を考え，いくつかの展開計算を効率よく行うことによって結果を得ている点である.

さらに，第２章の後半では，これらの結果を利用して T^2 型検定統計量を改良した新たな変換検定統計量をいくつか提案している．これらの変換検定統計量は，そのバイアスが T^2 型検定統計量のバイアスよりも縮小されたものになっている．具体的には理論的に変換検定統計量の期待値の N^{-1} の項の係数がゼロになっており，カイ二乗分布への近似がよいものになっている．また，モンテカルロ・シミュレーションとして，いくつかのパラメータに対して２ステップ単調欠測データとなるような欠測パターンの多次元正規乱数を発生させ，理論的結果の数値的評価を行っている.

第３章は，２標本問題における平均ベクトルの検定問題であり，分散共分散行列が異なる場合の検定を取り扱っている．一般に分散共分散行列が等しい場合はホテリングの T^2 統計量と呼ばれる検定統計量が通常用いられ，その帰無分布が F 分布で表現できることから F 分布の上側パーセント点を用いて検定を行うことができる．しかしながら，分散共分散行列が異なる場合は，多変量ベーレンス・フィッシャー問題と呼

ばれ，難しい問題として知られている．そのような背景の下で，本論文では，先行研究にある検定統計量とその帰無分布について再考察し，帰無分布に対する新たな近似分布として，自由度調整によるF分布を用いた近似分布を提案している．具体的には，検定統計量の期待値と分散をサンプル数に関して漸近展開することによって，F分布の自由度を求めるという方法である．特筆すべき点は，上記の漸近展開の計算に関して，N^{-2}の項までの漸近展開を与えることに成功している点である．これは，先行研究で与えられていたN^{-1}の項までの漸近展開を，多変量正規分布に従う確率ベクトルやウィシャート分布に従う確率行列の高次モーメントに関する期待値を与えることによって，N^{-2}の項まで拡張している．そして，これらの結果を基にF分布の自由度と照らし合わせて，自由度を調整したF分布による近似分布を提案している．最後に，モンテカルロ・シミュレーションを通して，先行研究による近似値と比較し，本論文で導出した近似がよい近似であることを数値的に示している．

以上のように，本論文は，1標本問題および2標本問題における平均ベクトルの検定について，それぞれ新たな検定統計量を与えるとともにその帰無仮説の下での近似分布を提案しており，多変量統計解析の理論に対する大きな貢献である．よって，本論文は学位（博士）論文として十分価値があると認める．