様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年　４月　２日現在

機関番号：17401
研究種目：基盤研究（C）
研究期間：2009 ～ 2011
課題番号：21540138
研究課題名（和文）　選択問題に対する数理統計的アプローチ

研究課題名（英文）　Statistical approaches to selection problems

研究代表者
高田　佳和（TAKADA YOSHIKAZU）
熊本大学・大学院自然科学研究科・教授
研究者番号：70114098

研究成果の概要（和文）：幾つかのグループから一つを何らかの基準に照らして選ばなければならない状況はよく起こる．この選択が、実験、調査等のデータにもとづいてなされるとき、最良母集団の選択という。この研究では、次の３つの選択問題に対する新たな手法を開発した。
1.　標準値、又はコントロールがある場合の最良母集団の選択
2.　データ間に相関がある場合の最良母集団の選択
3.　最良母集団の選択後、その母数の推定

研究成果の概要（英文）：We often encounter the case that we must select one of several groups based on a criterion. If the selection is based on the data obtained through experiments or investigations, it refers to a selection of the best population. In this research we provided three new procedures for the following problems
1.　The selection of the best population when there exists a standard or a control.
2.　The selection of the best population when there exists a correlation between data
3.　The estimation of the parameter after the selection of the best population

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009 年度	900,000	270,000	1,170,000
2010 年度	800,000	240,000	1,040,000
2011 年度	1,000,000	300,000	1,300,000
年度			
年度			
総　計	2,700,000	810,000	3,510,000

研究分野：統計学
科研費の分科・細目：数学・数学一般（含確率論・統計数学）
キーワード：最良母集団の選択、正規分布、サンプルサイズ、二段階法、三段階法

１．研究開始当初の背景

データを用いて、最良母集団を選択するときに問題になるのは､用いられた選択方法が、正しく最良母集団を選択しているかどうかであり、その確率が重要である。その確率を予め決められた値以上となる選択方法を構成したい。しかし、データ数を事前に固定するとそのような選択方法は存在しないことが分かっている。そのため、その問題を解くには、標本数の決定を途中で得られるデータにもとづいて決める逐次的手法が必要である。特に、選択問題では、二段階法（初期標本にもとづき､全本数を決める）がよく用いられている。色々な研究者がその手法を用いて、様々な選択問題について研究を行っている。しかし、その理論的特性については未知の部分も多く

あり、まだ解決されていない問題も多々ある。

２．研究の目的

この研究では、逐次手法の中で、特に二段階法と三段階法（第一、第二段階の標本を用いて、標本数を決める）にもとづき、以下に述べる選択問題の研究を行う。

(1) 標準値、対照がある場合の選択

例えば、新鋼材の開発において、その強度が、ある決められた値（既知）以上必要であることが求められる、また、新薬の開発において、その効果が市販されている薬の効果(未知）以上が求められる。このような場合の選択問題を、標準値、対照母集団がある場合の選択問題という。

(2)最良成分の選択

例えば、治療薬の比較で、同一の被験者に対して、時間をおいて異なる治療薬を投与し、その効果を比較する場合がある。この場合、得られたデータ間には通常相関が見られる。このような場合の選択問題を最良成分の選択という。

(3) 選択と推定

最良母集団の選択後、その母数を推定（長さ一定の信頼区間を用いて）したい場合がある。新たに、選択した母集団からデータを取って推定する場合は問題はないが、同じデータを用いて行う場合、選択方法の信頼性と信頼区間の信頼性を同時に満たす手法を開発する。

３．研究の方法

k 個の正規母集団 $N(\mu_1,\sigma_1^2),\cdots,N(\mu_k,\sigma_k^2)$ の母平均 $\mu_1,\cdots,\mu_k$ を大きさの順に並べ替えた値を $\mu_{[1]}\le\cdots,\le\mu_{[k]}$ としたとき、$\mu_{[k]}$ を母平均に持つ母集団が最良母集団とし、その選択問題を研究する。選択基準としては、Indifference zone approach を用いる。標本数の定め方としては、二段階法、または三段階法を用いる。二段階法としては、Stein (1954)、又は、Dudewicz and Dalal (1975) の方法を用いる。特に、後者の方法は、母分散が異なるとき、有効である。三段階法としては、Takada(2008)の修三段階法を用いる。

(1) 標準値、対照がある場合の選択

① 標準値がある場合の選択

μ_0（既知）を標準値とする。$\mu_{[k]}\le\mu_0$ ならばどの母集団も選択しないことが正しい選択(CS)となる。$\mu_{[k]}>\mu_0$ ならば、最良母集団を選択することが正しい選択（CS）となる。このとき CS の起こる確率を予め与えられた確率 $P^*\ (0<P^*<1)$ 以上にしたい。しかし、母数に制限を設けないと、条件を満たす標本数を決めることはできない。そこで、Bechhofer and Turnbull (1978)の Indifference zone approach を用いる。すなわち、$\delta^*(>0)$ を与え、$\mu_{[k]}\le\mu_0$ ならば、$P(CS)\ge P^*$ を $\mu_{[k]}>\max(\mu_0,\mu_{[k-1]})+\delta^*$ ならば $P(CS)\ge P^*$ を満たすように標本数を定める。

Bechhofer and Turnbull は、等分散（$\sigma_1^2=\cdots=\sigma_k^2$）の場合に確率要求を満たす標本数を二段階法を用いて求めている。その結果を分散が一般の場合にどの程度拡張できるかについて研究する

② 対照がある場合の選択

標準値がある場合の選択問題において、標準値の代わりに、対照正規母集団(その母平均 μ_0 は未知）を考える。上記の Bechhofer and Turnbull (1978)の Indifference zone approach を用いて標本数を定める。特に、等分散でない場合について研究する。

(2)最良成分の選択

多変量正規分布 $N_k(\mu,\Sigma)$ を仮定し、その平均ベクトル $\mu'=(\mu_1,\cdots,\mu_k)$ の最良成分を選択する。選択基準としては、Bechhofer (1954)の Indiference zone approach を用いる。すなわち、$\delta^*(>0)$ を与え、$\mu_{[k]}-\mu_{[k-1]}\ge\delta^*$ ならば、$P(CS)\ge P^*$ を満たすように標本数を定める。

分散・共分散行列 Σ が未知の場合、Clark and Yang (1986)が二段階法を用いて標本数を決定している。その標本数の特性を研究し、それを改良する新たな選択法を三段階法を適用して研究する。

(3) 選択と推定

k 個の正規母集団の中で最良母集団を選択し、かつ、その母数が長さ一定の信頼区間に含まれる確率を保証する標本数を求める。選択基準としては、Bechhofer (1954)の indiference zone approach を用いる。すなわち, $\delta^*(>0)$、$d\ (>0)$ を与え、$\mu_{[k]}-\mu_{[k-1]}\ge\delta^*$ ならば、$P(CS,|\hat{\mu}_{[k]}-\mu_{[k]}|<d)\ge P^*$ を満たす標本数を求める。ここで、$\hat{\mu}_{[k]}$ は $\mu_{[k]}$ の推定量である。等分散の場合とそうでない場合について二段階法を用いて研究する。等分散の場合は Stien の方法を、そうでない場合は Dudewicz and Dalal の方法を用いる。

４．研究成果

(1) 標準値、対照がある場合の選択

① 標準値がある場合の研究成果

・<u>分散既知</u>

等分散の場合の各母集団からの標本数を決める Bechhofer and Turnbull の公式において、分散を各母集団の分散に置き換えても、正しい選択をする確率が与えられた値以上になることを確かめた。さらに、その公式における色々定数の定め方によって決まる標本数の比

較を行った。

・分散未知

分散が未知の場合、Rinott (1978), Dudewicz and Dala, Lam (1988)の方法が適用できる。これら３つの方法の標本数を比較した。その結果、Dudewicz and Dala の方法が最適であることがわかった。しかし、Lam の方法との数値的な差は､それほど大きくない。簡便性から Lam の方法が推奨できる。

②対照がある場合の研究成果

Rinott と Dudewicz and Dala の方法を用いた標本数の定め方の研究を行った。さらに、２つの方法の標本数を比較した。その結果 Rinott の方法よりも Dudewicz and Dala の方法の方が標本数が少ないことが分かった。しかし、数値的評価では、その差はわずかであることも分かった。

③ 得られた成果と今後の展望

母分散が一般の場合の標本数の求め方については、その重要性にもかかわらず、これまであまり研究されていない。この研究で得られた成果は、これまで提案されている色々な方法の標本数の比較を行うことで、それぞれの方法の特性を明らかにすることができた。以下のことが今後の研究課題として考えられる。

・三段階法の応用

提案された手法は二段階法にもとづいている。二段階法には、不必要に標本数を多くとるという欠点がある（二次の漸近有効性が成り立たない)。その欠点を改良するため、三段階法の適用が考えられる。しかし、従来の三段階法は、二次の漸近有効性は成り立つが、確率要求は漸近的にしか成立しない。これを改善するために、Takada(2008)の修正三段階法の適用を考える。この方法では、確率要求を満たし､かつ二次の漸近有効性も成立する。二段階法に代わる方法として有効である。

・ロバスト性

等分散を仮定した場合とそうでない場合については､標本数の求め方が異なる｡しかし、他の問題では、等分散を仮定して求めた標本数の方が、そうでない場合の標本数より、母分散が大きく異ならなければ少なくてすむことが分かっている。このことがこの場合も成立するかどうかを研究する。すなわち、等分散からの乖離が CS にどのように影響するか、更に､一般の場合との標本数の比較を研究し、そのロバスト性を明らかにする。

(2) 最良成分の選択

① 研究成果

Clark and Yang (1986)が提案した手法の標本数の特性について研究を行った｡その結果、分散・共分散が既知な場合に必要となる標本数に比べて、その標本数は過度に大きくなることが分かった（二次の漸近有効性が成立しない）。このことを改良するために、takada(2008)の修正三段階法を適用して、標本数を定めた｡更に､理論的に Clark and Yang の方法よりも標本数が少なくなることを示した。その結果を確認するためにシミュレーションを行ったところ、理論的結果と同じ結果を示した。

② 得られた成果と今後の展望

Takada(2008)の修正三段階法は、二段階法の欠点を改良する方法として有効であることが示された。以下のことを今後の研究課題としている。

・分散・共分散が構造を持つ場合

Nelson and Matejcik (1995)は、分散・共分散行列Σがある種の構造を持つ時、Clark and Yang の方法を、標本数に関して改良する手法を提案している。しかし、修正三段階法を適用すれば、Nelson and Matejcik の方法を標本数に関し改良する手法の構成が可能と考えられる。

・標準値、対照成分ある場合の選択

標準値を超える成分があれば、その中で最良の成分を選択し、なければ、どの成分も選択しない問題を研究する。又、k 個の成分の一つを対照成分として、同様の問題を研究する。

・ロバスト性

分散・共分散行列Σに構造を仮定して構成した選択方法において、Σがその構造から乖離したときの選択方法のロバスト性について研究する。特に、実際の CS の確率が想定された値からどの程度ズレるか、及び、標本数の特性を明らかにする。

(3) 選択と推定

① 等分散の場合の研究成果

二段階法を用いて確率要求を満たす選択方法と推定方法の構成を行った。更に、その標本数の特性を漸近的な方法を用いて調べた。その結果、一次の漸近有効性を満たすことがわかった。共通の分散に関して既知の下限を仮定できれば、二次の漸近有効性を満たすこともわかった。その結果をシミュレーションで確認したところ、同じような結果が得られた。

② 分散が一般の場合の研究成果

等分散の場合は、選択方法として標本平均を用いることができるが、分散が一般の場合は、その問題は未解決である。しかし、標本平均を用いない Dudewicz and Dalal の方法を適用すれば、確率要求を満たす手法が構成できた。

③ 得られた成果と今後の展望

問題の重要性にも関わらず、これまでこの成果に類した結果は得られていなかった。この研究成果を踏まえて、以下のことを今後の研究課題としている。

・Confidence statement

選択に関しては、Bechhofer(1954) の Indifference zone approach を用いているので、母数が indifference zone すなわち、$\mu_{[k]} - \mu_{[k-1]} < \delta^*$ に属するとき、信頼性を表す確率 $P(CS, |\hat{\mu}_{[k]} - \mu_{[k]}| < d)$ については、何も分からない。それを改善するために、母数が indifference zone に属するときの信頼性を評価する必要がある。そのための方法が、Confidence statement である。この場合、$P(0 < \mu_{[k]} - \mu_S < \delta^*, |\hat{\mu}_S - \mu_S| < d) \ge P^*$ となる。ここで、μ_S は選ばれた母集団の母平均を表し、$\hat{\mu}_S$ はその推定量である。通常の最良母集団の選択に関しては、Indifference zone approach と Confidence statement は同等であるが、この場合はそれが成立しない。すなわち、Confidence statement が成立すれば、Indifference zone approach が成立するが、逆は成り立たない。Confidence statement が成立するような標本数の決定が今後の課題である。

・分散が一般の場合

Dudewicz and Dalal の方法を適用したが、より簡便な標本平均にもとづく Stein の方法の適用可能性を調べる。

・最良成分の選択と推定

多変量正規分布の母平均ベクトルの最良成分を選択した後に、その母数を推定する方法を研究する。

・三段階法の応用

等分散の場合、分散の下限が仮定できれば、標本数に関して、二次の漸近有効性が成立した。しかし、一般に適切な下限の設定が困難である。下限の設定が悪いと、過度に標本数を必要とする。この欠点を改良するために修正三段階法の適用が考えられる。

・ロバスト性

等分散を仮定した構成した選択方法において、等分散性から乖離したときのロバスト性について研究する。特に、実際の信頼性を表す確率が想定した値からどの程度ズレるか、及び、標本数の特性を明らかにする。

５．主な発表論文等

〔雑誌論文〕（計　５件）

① Yoshikazu Takada, Discussion on “Two-Stage Procedures for High-Dimentional Data” by Makoto Aosima and Kazuyoshi Yata, Sequential Analysis, 30, pp.400-403, 2011, 査読有

② Yoshikazu Takada, A three-stage selection procedure with an exact consistency, Communications in Statistics – Theory and Methods, 39, pp.1574-1584, 2010, 査読有

③ Yoshikazu Takada, Selecting the best normal population better than a standard under unequal variances, Journal of Statistical Planning and Inference, 140, pp.2693-2705, 2010, 査読有

④ Yoshikazu Takada, Selecting the best component of a Multivariate normal distribution, Communications in Statistics – Theory and Methods, 38, pp.3198-3212, 2009, 査読有

⑤ Makoto Aoshima and Yoshikazu Takada, Asymptotically optimal allocation for multiple comparisons with a control when variances are unknown and unequal, American Journal of Mathematical Management and Science, 29, pp.125-137, 2009, 査読有

〔学会発表〕（計　５件）

① 高田佳和，　最良正規母集団の選択とその母数の推定，研究集会“統計的推測の理論と方法論、及び、最近の動向”，2011,11,8, 筑波大学（茨城）

② 三山勝也，高田佳和，等分散を仮定した母平均の差の長さ一定の信頼区間の頑健性について，　統計関連学会連合大会，2011.9.6, 九州大学（福岡）

③ Yoshikazu Takada, Simultaneous Selection and Estimation of the Largest Normal Mean, The 58th Session of the International Statistical Institute, 2011,8,25, Dublin (Ireland)

④高田佳和、　標準値がある場合の最良正規母集団の選択-等分散が仮定できない場合、研究集会“統計的推測・確率解析とその周辺の話題の理論と応用”，　2009.12.12，秋田市カレッジプラザ

⑤ Yoshikazu Takada, Selecting the best normal population better than a standard under the unequal variance case, Second International Workshop in Sequential Methodologies, 2009,6,15, Troyes (France)

６．研究組織

(1)研究代表者

高田　佳和（TAKADA YOSHIKAZU）
熊本大学・大学院自然科学研究科・教授
研究者番号：70114098