

研究種目：基盤研究（C）  
 研究期間：2006～2008  
 課題番号：18500221  
 研究課題名（和文） 積分幾何および逐次解析に基づく多重比較法の展開とその  
 遺伝連鎖解析への応用  
 研究課題名（英文） Multiple comparisons procedures based on integral geometry and  
 sequential analysis and their applications to genetic linkage  
 analysis  
 研究代表者  
 栗木 哲 (KURIKI SATOSHI)  
 統計数理研究所・数理・推論研究系・教授  
 研究者番号：90195545

研究成果の概要：本研究は近年の多重比較法の理論的な進展をふまえて、多重比較法の理論展開とそれらの遺伝連鎖解析への応用を目指したものである。近年多重比較法は、逐次解析、積分幾何学的方法、組合せ数学・代数的方法によって理論的な発展が見られ、また応用の側面ではゲノムデータ解析の重要性が増大している。このような現状をふまえ、本研究では、遺伝連鎖解析への応用を念頭においた多重比較法の手法開発の研究を行い、以下の結果を得た：(i)  $d$ 次元上で定義され直積型の相関構造を持つカイ2乗確率場の、格子点上の最大値の上側裾確率の近似公式を、非線形再生理論の方法によって与えた。(ii) QTL解析の手法である区間マッピング法の、ロッドスコアの確率過程としての挙動を調べ、オイラー標数法によってその最大値を近似する方法について検討した。(iii) マウス2系統の行動データについてQTL解析と共分散構造解析を行い、行動を規定する遺伝子の連関についての仮説を得た。(iv) チューブ法の考え方を用いて、ガウス歪対称確率行列の最大特異値の寄与率分布を与えた。その結果を、誤差分散が未知の場合のSchefféの一对比較モデル、あるいは過分散を持つBradley-Terryモデルにおける三すくみ多重検定に応用した。

## 交付額

(金額単位：円)

	直接経費	間接経費	合計
18年度	1,300,000	0	1,300,000
19年度	1,100,000	330,000	1,430,000
20年度	1,100,000	330,000	1,430,000
年度			
年度			
総計	3,500,000	660,000	4,160,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：統計的推測

## 1. 研究開始当初の背景

多重比較法は、観測されたデータセットに基づいて、複数の仮説（命題）の真偽を判定する方法論である。単一の仮説の真偽を判定する統計的方法は、仮説検定（統計的仮説検定）と呼ばれる。多重比較法は、この仮説検定をいわば多重に行うことに相当し、このことから多重検

定と呼ばれることもある。実際のデータ解析では、一つのデータセットについて、興味ある（すなわちその真偽を判定したい）仮説が一つであるという状況はむしろ稀であり、多重比較法はデータ解析の現場からの要請に後押しされるかたちで比較的古い時期からその方法論が研究されてきた。

しかしながら従来研究において想定されてきた仮説の数は、数個から高々十数個である。一方で近年の大量データの解析、例えば染色体の全域にマーカーを配置した全ゲノムデータ解析やサンプリング間隔が非常に小さい時系列データでは、数百から数十万個の仮説を同時に取り扱う必要がある。また概念的には無限個の仮説を同時にとり扱う問題もある。これらの場合においては、従来の方法論がそのままの形で適用できるとは限らない。

このように多重比較法が対象とする領域が拡大する一方で、近年は理論的側面においても多重比較法の方法論一般の研究が進みつつある。その例としては、

- (a) 逐次解析・非線形再生理論を用いる方法、
  - (b) 積分幾何学的方法（オイラー標数法）、
  - (c) 組合せ数学・代数的方法（アブストラクトチューブ法）、
- などがある。

以上のような背景をふまえて、本申請研究においては、次項（研究の目的）に列挙する課題について取り組むことを通して、多重比較法の一般理論の展開、ならびに多重比較法の重要な応用分野であるゲノムデータ解析に焦点をあてた方法論の開発を行うことを目的とした。

## 2. 研究の目的

(i) 多次元の添字を持つ確率変数は確率場と呼ばれる。例えば $d$ 次元ユークリッド空間 $R^d$ の部分集合 $S$ を添字集合に持つ確率場を $X(t)$  ( $t \in S$ ) のように書く。確率場の構造（確率的挙動）が他と異なるような添字の値を、確率場の変化点という。例えば、期待値を $E[\cdot]$  で表すとき、 $t_0$ 以外の全ての点 $t$ で $E[t] \equiv \mu$  がなりたち、さらに $E[t_0] \neq \mu$  のとき、点 $t_0$ を変化点という。このような確率場の変化点を検出する手法は変化点解析とよばれ、多重比較法の一つと捉えることができる。この課題 (i) では、添字の次元 $d$ が2以上で、添字集合 $S$ が格子点集合でその格子間隔が密（結果として添字集合 $S$ の要素数が大）の場合を想定する。この状況は、遺伝連鎖解析への応用を念頭におくものである。逐次解析・非線形再生理論をツールとして、確率場変化点解析の判定閾値の設定、誤り確率（偽陽性、偽陰性）の評価などの一般理論とそれらの数値計算法の研究を目的とした。

(ii) 積分幾何学的方法（オイラー標数法）とは、添字集合 $S$ が連続集合である確率場 $X(t)$  ( $t \in S$ ) の最大値の分布の近似手法である。 $X(t)$ が $t$ の滑らかな関数であるなどの正則条件が満たされる場合は、オイラー標数法によって確率場の変化点解析における誤り確率を近似することができる。ここでも、遺伝連鎖解析に現れるいろいろな具体的な問題を念頭において、オイラー標数法近似公式の遺伝統計学への適用することを目的とした。

(iii) 計画的に交配されたイネやマウスのゲノムデータから、個体の形質を支配する遺伝子を特定する問題（QTL解析）や生殖的隔離障壁を特定する問題は、典型的な変化点解析である。(i)、(ii)に述べた方法を実ゲノムデータに適用し、その有効性を検証することも目的とした。

(iv) チューブ法、アブストラクトチューブ法の一般理論についてさらに発展させることも目的の一つであった。アブストラクトチューブ法は、Naiman and Wynn (1982), The Annals of Statistics, によって提唱された組合せ数学・代数的方法に基づく多重比較法である。いろいろな多様体や単体複体を添字集合に持つ確率場の場合、そのチューブ公式、アブストラクトチューブ公式を具体的に求めることは、容易でない場合が多い。いろいろな統

計モデルで必要とされるチューブ公式を具体的に与えることも目標とした。

## 3. 研究の方法

(i) 最初に、確率場の変化点解析の研究にむけて先行研究のサーベイを行った。D. Siegmund の学派によって、逐次解析・非線形再生理論に基づく確率過程（すなわち添字集合が1次元の確率場）の変化点解析の一般論ならびにその遺伝連鎖解析への適用に関する一連の論文が刊行されている。とくに W. Feller, M. Woodroofe, D. Siegmund らによる既存文献における数学技法は、本申請研究においても用いることができるものが多いため、文献サーベイを行った。

これらの準備をもとに、主として遺伝連鎖解析への応用を念頭において、確率場変化点解析の判定閾値の設定、誤り確率（偽陽性、偽陰性）の評価、サンプルサイズ設計（変化点検出のために必要なデータ数の見積もり）などを、できるだけ一般的な状況で理論展開を行った。

またここで用いる逐次解析・非線形再生理論は、いわゆる大偏差型の近似理論であり、データのサンプリング間隔を0に近づけたときの極限、あるいは確率分布の裾の挙動を論じる極限理論であるため、その近似が妥当であるかの吟味のための数値的検証を行った。

(ii) また平成17年度以前の研究実績をふまえて、積分幾何学的方法（オイラー標数法）の確率場の変化点解析、特に遺伝連鎖解析への適用を考えた。遺伝連鎖解析の分野で重要な QTL 解析（量的形質遺伝子座を特定するための解析）では、染色体上のマーカー数があまり多くない場合は、マーカーとマーカーの間を補完する区間マッピング法あるいはその派生法が用いられる。区間マッピング法による QTL 解析は、滑らかなサンプルパスを持つ確率過程、確率場の変化点解析と捉えることができるため、オイラー標数法が有効に働くことが予想されていた。

(iii) 国立遺伝学研究所による計画的に交配されたイネやマウスのゲノムデータに対して、(i)、(ii)で開発された方法論を適用する。

(iv) またそのことと平行して、チューブ法の一般理論の研究を行った。

## 4. 研究成果

(i)  $d$ 次元上で定義され直積型の相関構造を持つカイ2乗確率場の、格子点上の最大値の上側裾確率の近似公式を、非線形再生理論の方法によって与えた。ここで考えている漸近論は、裾確率の閾値 $b$ を無限大に、また格子点の間隔 $\Delta$ を0に近づけるものである。この近似公式をQTL解析における遺伝子座相互作用（エピスタシス）の検定へ適用しようとした際、問題となるのは上述の漸近近似による誤差、ならびにマーカー遺伝子間隔を等間隔 $\Delta$ とおくことによる近似誤差の2点である。これらについて、連鎖解析の典型的な例に即して、シミュレーションにより数値的な検討を行った。

(ii) QTL解析の代表的な手法である区間マッピング法の、ロッドスコアの確率過程としての挙動を調べ、オイラー標数法によってその最大値を近似する方法について検討した。比較的標準的な尤度比確率場の議論 (Ibragimov and Hasminskii, 1981) によって、サンプルサイズ $n$ を無限大とすると、ロッドスコアは滑らかなサンプルパスを持つカイ2乗確率場に分布収束することを示した。また

極限カイ2乗確率場の最大値分布の、オイラー標数法による近似を導いた。

(iii) マウス2系統の行動データについてQTL解析を行った。さらにその結果に基づいて共分散構造解析を行うことにより、行動を規定する遺伝子の連関についての仮説を得た。

(iv) チューブ法の考え方をを用いて、ガウス歪対称確率行列の最大特異値の寄与率分布を与えた。その最大特異値の分布は文献(栗木, 1991, 応用統計学)などによって既に知られていた。最大特異値がある多様体上のガウス確率場の最大値とみなすことができ、一方で最大特異値の寄与率分布が球面チューブの体積に対応づけることができることから、チューブ法の考え方を逆に辿ることによって、最大特異値の分布から寄与率の分布を導いた。また導出した分布を、誤差分散が未知の場合のSchefféの対比較モデル、あるいは過分散を持つBradley-Terryモデルにおける三すくみ多重検定に応用した。

(v) 2006年11月27日(月), 28日(火)に研究集会「統計モデルの数理と実際」(統計数理研究所)を科学研究費基盤研究(A)18200019(研究代表者:竹村彰通)との共催で開催し、遺伝連鎖解析などについて研究打ち合わせを行った。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計5件)

①Kuriki Satoshi, Takemura Akimichi, Euler characteristic heuristic for approximating the distribution of the largest eigenvalue of an orthogonally invariant random matrix, *Journal of Statistical Planning and Inference*, Vol. 138, Issue 11, 3357-3378 (2008).

②Kuriki Satoshi, Takemura Akimichi, The tube method for the moment index in projection pursuit, *Journal of Statistical Planning and Inference*, Vol. 138, Issue 9, 2749-2762 (2008).

③栗木哲, 竹村彰通, チューブの体積と正規確率場の最大値の分布, *数学*, 60巻, 2号, 134-155 (2008).

④Oka Ayako, Aoto T., Totsuka Y., Takahashi R., Ueda M., Mita A., Sakurai-Yamatani N., Yamamoto H., Kuriki Satoshi, Takagi N., Moriwaki K. and Shiroishi T., Disruption of genetic interaction between two autosomal regions and the X chromosome causes reproductive isolation between mouse strains derived from different subspecies. *Genetics*, Vol. 175, Issue 1, 185-197 (2007).

⑤Gao Wei and Kuriki Satoshi, Testing marginal homogeneity against stochastically ordered marginals for  $r \times r$  contingency tables. *Journal of Multivariate Analysis*, Vol. 97, Issue 6, 1330-1341 (2006).

[学会発表] (計13件)

①Kuriki Satoshi, Harushima, Y., Fujisawa, H. and Kurata, N., Multiplicity Adjustments in Detecting Reproductive Barriers caused by Loci Interactions, BIRS Workshop 09w5040 Random Fields and Stochastic Geometry, Banff (Canada), 2009.2.26

②栗木哲, Distributions of the largest singular values of skew-symmetric random matrices and their applications to paired comparisons, *Recent Advances in Statistical Inference --- in Honor of Professor Masafumi Akahira*, 筑波大学, 2008.12.16

③Kuriki Satoshi and Anthony Hayter, Testing superiority in polynomial regressions, *Computational Algebraic Statistics, Theories and Applications (CASTA2008)*, 京都大学, 2008.12.10

④栗木哲, 歪対称確率行列の最大特異値の寄与率分布とその対比較への応用, 統計関連学会連合大会, 慶応義塾大学(横浜), 2008.9.9

⑤Kuriki Satoshi, The tube method for the moment index in projection pursuit, *The 2nd Joint Meeting of ISI, ISM, and ISSAS, Academia Sinica (台北)*, 2008.6.19

⑥Kuriki Satoshi, Multiplicity Adjustments in Detecting Reproductive Barriers caused by Loci Interactions, *First joint meeting between Institute of Statistical Science, Academia Sinica, Taiwan and the Institute of Statistical Mathematics, 統計数理研究所*, 2007.11.29

⑦栗木哲, 格子点上カイ2乗確率場の最大値分布の近似とその連鎖解析への応用, 科研費研究集会「統計的モデリングの方法と理論」, 一橋大学, 2007.11.27

⑧栗木哲, 生殖的隔離に関わるエピスタシス(相互作用)検出と多重性調整, 融合研究シンポジウム「地球と生命の新パラダイム創造への挑戦」(ハイライト講演), 学術情報センター, 2007.10.18

⑨栗木哲, 遺伝子座間の相互作用による生殖的隔離障壁の検出と多重性調整, 統計関連学会連合大会, 神戸大学, 2007.9.6

⑩栗木哲, 直積型の相関構造を持つカイ2乗確率場の最大値の分布, 日本数学会2007年度年会, 埼玉大学, 2007.3.29

⑪栗木哲, 非心ウィシャート分布のモーメントとラグエル多項式, 確率解析と統計学, 統計数理研究所, 2007.2.15

⑫栗木哲, オイラー標数法とアブストラクトチューブ法, 科学研究費研究集会「計算代数統計の展開」, 京都大学, 2006.11.8

⑬栗木哲, 積分幾何と統計分布理論, 大阪市立大学数学研究所ミニスクール「情報幾何への入門と応用」(招待講演), 大阪市立大学, 2006.6.9~10

[図書] (計1件)

①栗木哲 「QTL解析の統計モデルと検定の多重性調整」, 21世紀の統計科学, II (小西貞則, 国友直人 編), 東京大学出版会, 315-356, 2008 (2008.7)

[産業財産権]

○出願状況 (計0件)

○取得状況 (計0件)

[その他]

6. 研究組織

(1) 研究代表者

栗木 哲 (KURIKI SATOSHI)

統計数理研究所・数理・推論研究系・教授

研究者番号：90195545

(2) 研究分担者

(3) 連携研究者