

## 科学研究費助成事業 研究成果報告書

平成 26 年 5 月 27 日現在

機関番号：17102

研究種目：基盤研究(B)

研究期間：2009～2013

課題番号：21340026

研究課題名(和文) 高度な情報抽出のためのノンパラメトリック推測理論の導出とその応用

研究課題名(英文) Improvement Theory of Nonparametric Statistical Precise Inference and Its Applications

研究代表者

前園 宜彦 (Maesono, Yoshihiko)

九州大学・数理(科)学研究科(研究院)・教授

研究者番号：30173701

交付決定額(研究期間全体)：(直接経費) 11,200,000円、(間接経費) 3,360,000円

研究成果の概要(和文)：データを繰り返し利用する統計的リサンプリング法の高精度近似理論を構築し、その性質を明らかにした。また滑らかな推定結果を与えるカーネル型推定量の分布の、精度の高い近似を求め、その結果をデータの順位を利用する検定へ応用する新しい手法を開発した。さらに複雑な非線形構造をもつ現象をガウス型の指数関数に基づいて推測する新しい手法を開発した。同時に空間に依存する性質をもつデータに対して、柔軟な非線形の回帰モデルを考案し、従来のモデルより推測の精度が改善されることを示した。

研究成果の概要(英文)：In this project, we obtain an higher order asymptotic properties for statistical inference based on resampling methods, which use the data repeatedly. We also obtain approximations of distribution functions of kernel type estimators, and applying these results, we propose new nonparametric test statistics. Furthermore, we develop an inference method based on Gaussian basic functions, which we can apply to non-linear complex systems. The non-linear regression model is studied precisely, and we obtain a flexible method which improve accuracy of statistical inferences for state space data.

研究分野：数物系科学

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：ノンパラメトリック カーネル型推定量 情報量規準 エッジワース展開 非線形判別 ブートストラップ法 ジャックナイフ法 リサンプリング法

## 1. 研究開始当初の背景

(1) ノンパラメトリックな統計的推測の研究は、順位検定、順位に基づく点推定・区間推定などさまざまな推測法が提案され、標本数を大きくしたときの中心極限定理に基づく理論的研究は一応の完成を見た。現在研究されている漸近理論は単純な中心極限定理から深化され、エッジワース展開、サドルポイント近似、正規化変換などがノンパラメトリックな設定の下で求められている。しかしこのような最新の成果の実用化はまだ程遠い状況である。生存解析の分野では Cox (J. Roy. Stat. Soc. 1972 など) によってコックスモデルと呼ばれるノンパラメトリックな手法が提案され、現在では生存時間解析やリスク解析の基本的な道具として定着している。しかしこのモデルの高次の漸近理論の構築はまだ研究途上で、高次理論の実用化はほとんどなされていない。

(2) 1980 年代に隆盛したりサンプリング法、特に Efron (Ann. Stat. 1979) によって提案されたブートストラップ法はコンピュータの発展に伴ってその有用性が高まり、ノンパラメトリック推測の中心的手法として現在も盛んに研究されている。しかしリサンプリング法は数学的根拠を理解するのが難しいために、実務の面では適切に利用されていないのが現状である。また統計的リサンプリング法は基本的な統計モデルについての性質は明らかになっているが、複雑な現実のデータに適用するときの問題点の克服はこれからである。

## 2. 研究の目的

(1) 現在までに得られている順位に基づく統計的推測法や、分布を特定せずに提案されている統計量の性質を利用して現実のデータの解析に適用できるように改良し、その理論的な性質を明らかにすることを目指す。具体的には順位に基づく点推定量の高次の漸近的性質を明らかにし、その成果を元に高精度の信頼区間の構成法を求め、また現実のデータに対する適用を容易にするアルゴリズムやプログラムの開発を行う。一方パラメトリックな設定の下で導入された標本平均、標本分散、標本相関係数などの基本統計量に対して、誤差項の従う分布が当初の仮定と異なるときにも有効に活用できる漸近理論の構築を図る。それらの研究成果を元に実用化の問題点を明らかにしその克服に取り組む。

(2) 近年のノンパラメトリック推測法の中心となっている統計的リサンプリング法の有効利用の研究を行う。ブートストラップ法に対してはリサンプリングの最適な繰り返し数についての研究を行い、高次の漸近理論と精度が同等な方法を求める。またジャックナイフ法とブートストラップ法の組み合わせについても研究し、最適な組み合わせ方やリサンプリングの適用順序を入れ替えた時の影響を明らかにする。さらにカーネル法やノ

ンパラメトリック回帰の数学的な性質を明らかにし、実用化のための問題解決を目指す。

## 3. 研究の方法

(1) ノンパラメトリック推測の研究者を集めてシンポジウムを開催する。この研究集会で現在のノンパラメトリック推測研究の現状、問題点及び今後の研究の方向性について参加者と議論し、本研究に関連する情報を収集する。具体的には以下の役割分担をもとに研究を進める。まず代表者・前園は、統計的推測におけるジャックナイフ法やブートストラップ法等のリサンプリング法の成果を、確率微分方程式に基づいた数理モデルや Huber 流の M-推定量に基づく信頼区間の構成に対して有効に適用するための改良を行なうとともに、新しい手法の開発を目指す。また統計的漸近理論の結果を利用して正規分布での近似の精度を上げる統計量を構成し、リサンプリング法と組み合わせる有効な推測法を開発する。同時に推測の精度を上げることが分かっている反復ブートストラップ法を、大量のデータに対して適用するための新しい手法である重み付きブートストラップ法の有効性を吟味し、これまで得られていた単純な統計モデルからより複雑なモデルに適用できるように手法の改良を目指す。さらに最新の統計手法であるカーネル法を利用して金融データ解析におけるバリュー・アット・リスクや地震リスクのための推定法の改善を図る。

(2) 分担者・小西は、多変量解析におけるリサンプリング法の問題点、特に反復ブートストラップ法を適用するときの効率の良い方法の開発を目指すと同時に数学的な性質を明らかにする。分担者・西井は画像解析の分野でのノンパラメトリック法の研究を推進していく。特に共同研究者の前園や小西と協力して統計的リサンプリング法を利用したブースティングの性質の解明を図る。分担者・寒河江はカーネル法に基づく密度関数やヒストグラムに基づく密度関数の推定法、ノンパラメトリックなデータ圧縮法などノンパラメトリックな設定の下での研究を行う。またこれまで主催してきたノンパラメトリック・セミパラメトリックの研究集会の蓄積を元に、21 年度にシンポジウムを開催し、本研究を推進するための情報収集及び意見交換を行う機会を作る。分担者・藤井は生物統計におけるコックスモデルの高次漸近理論の研究を行う。また分担者・二宮は変化点解析におけるノンパラメトリック推測の有効な適用法を研究する。

(3) 上記の役割分担のもとで研究を遂行するために、高性能のパソコンを購入して、計算機ソフトによる研究成果の実用性のチェックを行なうと同時に、大学院生の支援を受けて計算機上でのシミュレーションを効率良く行う。シミュレーションのためのソフト及び統計ソフトを導入し、得られた理論の検証

を行ない、実用的なプログラムの開発を目指す。またこれまでに得られているブートストラップ法の研究成果をシミュレーションで検証し、統計的リサンプリング法の新しい応用の可能性を探る。さらにより詳しく研究成果の位置づけ及び研究の方向性についてアドバイスを求めるために、代表者・前園がメルボルン大学の Peter Hall 教授を訪問し研究のレビューを受ける。

#### 4. 研究成果

(1) ノンパラメトリックな統計的推測における高次モーメントのジャックナイフ推定量の漸近的な性質を求め、その有効な利用法について検証し、ブートストラップ分散推定量の漸近表現と漸近平均二乗誤差を求めることができた。またスチューデント化カーネル型確率点推定量のエッジワース展開を求め、理論的に妥当性を証明するとともに、この結果を使って確率点推定量に基づくノンパラメトリックな信頼区間を構成した。さらに  $U$ -統計量に対する統計的リサンプリング法の有効性を、高次の漸近理論を利用してシミュレーションだけではなく、理論的性質も明らかにした。スチューデント化確率点推定量の高次漸近分布を求め、それを利用して精度の高い近似信頼区間の構成に成功した。また確率密度関数のカーネル型推定量の正規近似の改良を求めてその数学的な妥当性を証明した。これらの成果は世界に先駆けてのものである。分布関数のカーネル型推定量の正規近似の改良を求めて、その有効性をシミュレーションで示すとともに、その数学的な妥当性についても示すことができた。この成果を利用して順位検定統計量の連続化を提案し、高次の正規近似を求めた。さらに経験分布関数の連続化であるカーネル型分布関数推定量を利用した符号検定の連続化の漸近展開について細かく分類することに成功し、その展開を具体的な形で求めた。求めた成果は3次の漸近理論に相当するもので、初めて得られたものである。

(2) 大規模モデルの推定と変数選択の新しい手法である lasso 推定を拡張して、非線形現象のモデル化に適用する研究に取り組み、新たな非線形モデリング手法を提唱した。モデル化の過程で、ベイズ推論によるアルゴリズムを組み込むことによって、柔軟なモデリング手法を構成することができた。複雑な非線形構造を内在する現象を捉えるモデルの構築を目的として、微分可能でない  $L_1$  タイプの正則化法に基づく回帰モデリングの開発研究に取り組んだ。その結果、ガウス型基底関数に基づく新たな非線形回帰モデリングを提案し、手法の有効性を立証した。これによって、滑らかに変化する現象の真の構造だけでなく、何らかの要因によって急激に変化している現象でも適切に捉えるモデルが構築できた。スパース回帰モデリングの研究及び正則化項の制約を制御するパラ

メータの選択に関するモデル評価基準を解析的に導出するとともに、適用上有効なブートストラップ計算アルゴリズムに基づく方法を提唱した。

(3) 地表面メッシュを被覆しているカテゴリ割合を推定する問題について、マルコフ確率場に従う正規混合分布によるアプローチを提案した。また森林被覆率の平均構造を自然3次スプラインで推定し、細かい変動を抽出することに成功した。同時に森林被覆率を人口密度と起伏量で説明する空間依存性をもつ柔軟な非線形回帰モデルを考案し、従来のパラメトリックな平均構造で記述するモデルより推測の精度を改善することを示した。さらに森林被覆率をゼロインフレーションおよび空間依存する logistic-normal 回帰分析により予測するモデルを提案した。

制約のある説明変数を用いた多項式回帰モデルにおいて、高精度のモデル推定法を提案し、車体設計の実験回数を大幅に減少させた。また時系列モデルのうち自己回帰成分を持つ回帰モデルについて、予測域の重要度に重みをつけたモデル選択基準を導出した。さらに全地球の弱い地震頻度に対する太陽風の影響を、ポアソン回帰や負の2項分布に基づく一般化線形モデルで検出した。

(4) 高次漸近特性を持つ多次元ノンパラメトリック確率密度関数の推定量を提案し、その有効性を示した。カーネル密度推定法を用いた非線形判別法を提案し、その特徴と欠点を明らかにした。また主成分分析、非負値行列因子分解、ベクトル量子化による次元縮小化後のカーネル密度推定法の有効性を検証した。ヒストグラムのインターバルを推定するモーメント法とその探索アルゴリズムを提案した。またヒストグラム密度関数のベイズ推定法と超パラメータの選択法を提案した。さらにピン化カーネル密度推定と  $k$  平均法による複数の正解分布を用いることで非線形判別の性能が向上することを示した。応用として、ベンチマーク用の手書き数字データの判別では、サポートベクタマシンによる判別とほぼ同等な性能を達成し、ノイズのあるデータでは本提案法が優れていることを示した。

(5) コホート内患者対照研究における対象の抽出デザインについて、無作為抽出とカウンターマッチングデザインによる曝露効果や交互作用項に対する検定の検出力の計算方法を比較検討した。また尤度比検定に対する検出力の計算方法に関して漸近的な方法を提案し、計算結果とシミュレーションによりその有効性を明らかにした。観察研究における検出力の計算と症例数の決定法について研究を進め、コホート内症例対照研究及びケースコホート研究における検出力の計算や比例ハザードモデルを仮定しない方法について新たな提案を行った。またペアワイズ条件付き尤度関数を用いた推定方法を考察し、それぞれの場面における適用方法の改善を行った。

(6) 空間データにおける集積性を検知するための尤度比検定において、 $p$  値のタイトな上限を容易な計算で与える方法を構築した。そして数値実験により、その有用性を示した。また集積性検出において確率値の上界を容易に与える新しい方法を導き、ノンパラメトリック回帰問題において滑らかさが変動する場合の新しいモデリング手法を考案した。現象過程、動作過程の高頻度観測データを関数化処理した関数化データ集合の中から高効率に情報やパターンを抽出するための理論・方法論の研究を行い、関数データ集合に基づく多変量重回帰モデリング手法を提唱した。離散観測される正規確率場の超過確率に対し、一般的な設定で上界を与える評価式を導出し、それをゲノム解析における QTL 検出問題に適用した。また脳画像解析のために、関数時系列データの構造変化を検出するための  $C_p$  規準を導出した。

#### 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 35 件)

Zhong Huang and Yoshihiko Maesono, Edgeworth Expansion for Kernel Estimators of a distribution function, Bulletin of Informatics and Cybernetics, 掲載決定

Yoshihiko Maesono and Spiridon Penev, Improved confidence intervals for quantiles, Annals of the Institute of Statistical Mathematics, Vol. 65, 2013, pp.167-189  
DOI 10.1007/s10463-012-0369-6

Shota Umeno and Yoshihiko Maesono, Improvement of Normal Approximation for Kernel Density Estimator, Bulletin of Informatics and Cybernetics, 査読有、Vol.45, 2013, pp.11-24

Kei Hirose and Sadanori Konishi, Variable selection via the weighted group lasso for factor analysis models, Canadian Journal of Statistics, 査読有、Vol. 40, 2012, pp.345-361,  
DOI: 10.1002/cjs.11129

Masayuki Mizuno and Yoshihiko Maesono, Mean squared errors of bootstrap variance estimators for U-statistics, Bulletin of Informatics and Cybernetics, 査読有、Vol.43, 2011, pp.67-82

Hidetoshi Matsui, Takamitsu Araki and Sadanori Konishi, Multiclass functional discriminant analysis and its application to gesture recognition, Journal of Classification, 査読有、Vol.28, 2011, pp.227-243,  
DOI:10.1007/s00357-011-9082-z

Yoshihiko Maesono, The jackknife, International Encyclopedia of Statistical

Science, Ed. Miodrag Lovric, Springer, 査読無、2011, pp.697--699

Shizue Izumi and Yoshinori Fujii, Power Calculation for likelihood ratio tests in the nested case-control designs with a randomly sampled control per case, 計量生物学, 査読有、Vol.31, 2010, pp.1-11

二宮嘉行, 吉本敦, ベイズ法を用いた単木成長予測モデル, FORMATH, 査読有、Vol.10, 2011, pp.333-349

前園宜彦, 高次モーメントのノンパラメトリック推定, 日本統計学会誌, 査読有、39 巻, pp.355-367, 2009,

<http://www.terrapub.co.jp/journals/jjssj/frame/39.html>

Yoshihiko Maesono, Edgeworth Expansion and Normalizing Transformation of Ratio Statistics and their Application, Communications in Statistics-Theory and Methods, 査読有、Vol.39, 2010, pp.1344-1358, 2010,

DOI:10.1080/03610920802311741

Yoshihiko Maesono and Spiridon Penev, Edgeworth expansion for the kernel quantile estimator, Annals of the Institute of Statistical Mathematics, 査読有、Vol.63, 2011, pp.617-644,  
DOI:10.1007/s10463-009-0241-5

Yuko Araki, Sadanori Konishi, Shuichi Kawano and Hidetoshi Matsui, Functional regression modeling via regularized Gaussian basis expansions, Annals of the Institute of Statistical Mathematics, 査読有、Vol.61, 2009, 811-833,  
DOI:10.1007/s10463-007-0161-1

Shohei Tateishi, Hidetoshi Matsui and Sadanori Konishi, Nonlinear regression modeling via the lasso-type regularization, Journal of Statistical Planning and Inference, 査読有、Vol.140, 2010, pp.1125-1134,  
DOI:10.1016/j.jspi.2009.10.015

Shojiro, Tanaka and Ruei, Nishii, Non-linear regression models to identify functional forms of deforestation in East Asia, IEEE Transactions on Geoscience and Remote Sensing, 査読有、Vol.47, No.8, 2009, pp.2617-2626,  
DOI:10.1109/TGRS.2009.2015659

寒河江 雅彦, 野呂 拓矢, デビッド・スコット, Polynomial Histogram による多次元ノンパラメトリック確率密度推定, 日本統計学会誌, 査読有、39 号、2 号、2009, pp.265-298,  
<http://www.terrapub.co.jp/journals/jjssj/frame/39.html>

[学会発表](計 20 件)

前園宜彦, 分布関数のカーネル型推定量の高次漸近分布とその応用、研究集会ノンパ

ラメトリック統計解析とベイズ統計、2014年3月19日、慶応大学

前園宜彦、符号検定の連続化と有意確率の近似について、日本数学会総合分科会、2013年9月27日、愛媛大学

前園宜彦、分布関数のカーネル型推定量の高次漸近分布、統計関連学会連合大会、2013年9月11日、大阪大学

M. Lu and Y. Maesono、Smoothing of sign test and approximation of its p-value、研究集会 Bernoulli Society Satellite Meeting to the ISI World Statistics Congress 2013、"Asymptotic Statistics and Related Topics: Theories and Methodologies"、2013年9月3日、東京大学

R. Nishii、P. Qin and D. Uchi、Nonlinear Bayesian unmixing of geospatial data based on Gibbs sampling、IEEE IGARSS 2013、2013年7月21日~7月26日、Melbourne Convention and Exhibition Centre

P. Qin、K. Yanase and R. Nishii、Multi-hour-ahead prediction of Dst index using autoregressive models with exogenous variables、Japan Geoscience Union Meeting 2013、2013年5月19日~5月24日、幕張メッセ

Yoshiyuki Ninomiya、P-value evaluation for multiple testing of means under the existence of positive correlations、8th International Conference on Multiple Comparison Procedures、2013年7月8日~11日、Southampton

前園宜彦、魯夢欣、符号検定の平滑化と有意確率の近似について、日本数学会年会、2013年3月22日、京都大学

前園宜彦、Spiridon Penev、Asymptotic distributions of kernel type estimators、IMS-APRM2012 サテライトシンポジウム、2012年7月5日(招待講演) 早稲田大学

松井秀、俊三角 俊裕、小西 貞則、L1 正則化に基づく変化係数モデリングと変数選択、統計関連学会連合大会、2012年9月11日、北海道大学

山本けい子、寒河江雅彦、カーネル密度推定法による非線形判別への応用、2012年度統計関連学会連合大会、2012年9月11日、北海道大学

Y. Fujii and S. Izumi、Design and method of the power calculations for likelihood ratio tests in the nested case-control studies、International biometrics conference、2012年8月、神戸

金大柱、川野秀一、二宮嘉行、局所適応型正則化法に基づく非線形回帰モデリング、統計関連学会連合大会、2012年9月10日、北海道大学

梅野翔太、前園宜彦、密度関数に対するカーネル型推定量の正規近似の改良、日本数学会総合分科会、2011年9月29日、信州大学

松田和己、保科架風、小西貞則、関連ベクターマシンに基づく非線形構造探索、統計関連学会連合大会、2011年9月6日、九州大学

宮田大樹、西井龍映、田中章司郎、森林被覆率の非線形回帰モデリング、統計関連学会連合大会、2011年9月6日、九州大学

二宮嘉行、因子分析における因子数選択のための分布理論、統計関連学会連合大会、2011年9月6日、九州大学

立石正平、小西貞則、Fused lassoに基づく非線形回帰モデリング、統計関連学会連合大会、2010年9月6日、早稲田大学

前園宜彦、高次モーメントのノンパラメトリック推測、日本数学会年会、2010年3月26日、慶応大学

Y. Maesono and S. Penev、Edgeworth expansion for a studentized kernel quantile estimator、日本数学会総合分科会、2009年9月26日、大阪大学

〔図書〕(計 1件)

小西貞則、多変量解析入門、岩波書店、2010、306ページ

〔産業財産権〕

出願状況(計 0件)

名称:

発明者:

権利者:

種類:

番号:

出願年月日:

国内外の別:

取得状況(計 1件)

名称: 車両用運転支援装置

発明者: 西井 龍映、秦 攀、新部 忠幸、  
中本 尊元、楠本 信平

権利者: 九州大学、マツダ株式会社

種類: 特許公開

番号: 2013-025521

取得年月日: 2013年02月04日

国内外の別: 国内

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

前園 宜彦 (MAESONO, Yoshihiko)

九州大学・大学院数理学研究院・教授

研究者番号: 30173701

(2) 研究分担者

西井 龍映 (NISHII, Ryuei)

九州大学・マス・ファオ・インダストリ研究所・教授

研究者番号： 40127684

(3)研究分担者

小西 貞則 (KONISHI, Sadanori)  
中央大学・理工学部・教授  
研究者番号：40090550

(4)研究分担者

寒河江 雅彦 (SAGAE, Masahiko)  
金沢大学・経済学経営学系・教授  
研究者番号：20125669

(5)研究分担者

藤井 良宜 (FUJII, Yoshinori)  
宮崎大学・教育文化学部・教授  
研究者番号：10218985

(6)研究分担者

小西 貞則 (NINOMIYA, Yoshiyuki)  
九州大学・マス・ファオ・インダストリ研  
究所・准教授  
研究者番号：50343330