

平成 23 年 3 月 31 日現在

機関番号：12102

研究種目：基盤研究(B)

研究期間：2007～2010

課題番号：19340020

研究課題名(和文) 生物情報を解明するための統計数学的基礎理論とその応用

研究課題名(英文) Theory of the foundation of mathematical statistics to analyze the biological information and its applications

研究代表者 赤平 昌文 (AKAHIRA MASAFUMI)

筑波大学・副学長

研究者番号：70017424

研究成果の概要(和文)：バイオインフォマティクス、医学分野等の情報解析における統計数学的基礎理論の研究を行った。特に、逐次選択実験問題について考察し、臨床試験における 2 つの処置法の適用計画において期待効果が最大になるような手法を開発した。また、2 標本問題における推測において重要な役割を果たす統計量が非心分布に従うことが多く、特に奇数自由度を持つ非心カイ 2 乗分布のパーセント点の近似式を解析的に導出し、カイ 2 乗検定の検出力と非心度の観点からも非心カイ 2 乗分布の近似の良さを数値的に評価した。さらに正規性の条件を課さないで、非心 t 統計量の分布のパーセント点の高次近似式を導出し、様々な分布の下でその近似式の精確性を数値的に確認した。

研究成果の概要(英文)：The research on basic theory of statistical mathematics in the information analysis of bioinformatics, medical sciences, etc. In particular, we consider the problem on sequential experiment of choice, and develop a procedure which maximizes the expected effect in an application plan of 2 treatments in clinical trials. The statistic playing an important role in the inference of two-sample problem follows a non-central distribution in many cases. We analytically derive an approximation formula of a percentage point of the non-central chi-square distribution with the odd degree of freedom, and numerically evaluate the goodness of the approximation of the non-central distribution from the viewpoint of the power and the non-centrality. Further, we derive a higher order approximation formula of the percentage point of the non-central t-statistic, and numerically verify the accuracy of the formula under various distributions.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007 年度	3,800,000	1,140,000	4,940,000
2008 年度	3,500,000	1,050,000	4,550,000
2009 年度	3,500,000	1,050,000	4,550,000
2010 年度	3,500,000	1,050,000	4,550,000
年度			
総計	14,300,000	4,290,000	18,590,000

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：生物統計、臨床統計、情報量、漸近有効性、一致性、非心分布、Bernoulli 分布、Cornish-Fisher 展開

1. 研究開始当初の背景

数理統計学の応用分野として医薬学分野があることはよく知られ、統計的手法が重要な役割を果たすことが認識されていた。特に、バイオインフォマティクス等の情報解析において統計解析がますます重要になってきていた。そのような中で、様々な生物情報に関して、統計解析の観点から多重決定問題、逐次選択実験問題、遺伝的情報を含む生物配列の解析等について、その解決に向けて統計数学的基礎理論の構築とその応用が非常に期待されていた。

2. 研究の目的

バイオインフォマティクス、医薬学分野等の情報解析における理論的基礎として統計数学がますます重要になってきている。たとえば、通常の統計的検定においては仮説を棄却、受容の2通りの決定を下すため、データが臨界値の近傍に落ちた場合の判断、あるいはデータが外れ値を含む場合の過誤の問題等が起こり、いくつかの考えられる決定の中から適切な1つを選ぶ多重決定理論はその問題を解決する方法の1つで、判断の誤りは極力避けねばならないような医薬学分野での応用に極めて有用になる。そこで、本研究では、多重決定問題について適当な損失関数を導入して適切な信頼方式を構成し、その応用についても考える。また、逐次選択実験において漸近的に有効な割り当て方式を求め、その逐次臨床試行における処理の割り当てへの応用についても考える。次に、非正則推測において、独自の情報不等式を導出して、それによる最適な推測方式を求め、理論的見地のみならず、数値的見地からもその評価について検討する。さらに、生物情報の解析については、統計的解析の観点からはまだ未開拓な部分が多く、統計解析の理論的基礎の構築とその応用を目指す。

3. 研究の方法

本研究を進めるためには、従来の統計的推測における推定、検定の他に、逐次選択実験的な方法を加味して行うことが必要であり、それは従来、必ずしも十分に検討されていない方法であった。また、分布の高次近似法を適用するために、特に2標本問題における非心分布の近似においては、Cornish-Fisher展開を用いてパーセント点の近似式を導出した。さらに、切断分布を含む非正則分布族の母数の推測においては、従来の正則な場合に有用なFisher情報量は使えないので、一般化情報量を用いてEdgeworth展開を駆使して統計量の漸近情報量損失を求めた。その他にも、高次の大偏差近似も統計量の分布の裾確率を計算するのに有用であった。

4. 研究成果

本研究課題の下で、バイオインフォマティクス、医薬学分野等の情報解析における統計数学的基礎理論の研究を行った。まず、逐次選択実験問題について考察した。臨床試験において2つの処置法の適用計画において、 N 人の対象者から n 人を選んで1つの処置法を適用し、その結果によって残りの $(N-n)$ 人に対する処置法を選択する。このとき、問題として、「 $(i)_n$ をどのように選ぶか、そして(ii)臨床試験の仕方をどのように決めるか」が考えられ、期待効果が最大になるような手法の開発に努めた。実際には、同じ成功確率をもつBernoulli分布に従う2つ独立な確率変数列を考え、ミニマックスルールの構築を試みた。

また、2標本問題における推定、検定において重要な役割を果たす統計量が非心分布に従うことが多いが、そのパーセント点を解析的に求めることは必ずしも容易でない。そこで、奇数自由度をもつ非心カイ2乗分布のパーセント点の近似式を解析的に導出した。そして、カイ2乗検定の検出力と非心度の観点からも非心カイ2乗分布の近似の良さを数値的に評価した。さらに、従来は正規性の仮定の下で非心 t 統計量が区間推定、検定で重要な役割を果たしてきたが、正規性の仮定をしない場合の非心 t 統計量の分布はほとんど論じられてこなかった。一般には、母集団分布が正規分布であるという保証はなく、非正規性の下で非心 t 統計量に基づく推測は極めて重要である。本研究において、独立同分布で6次のモーメントをもつ非退化連続型確率変数列の場合に非心 t 統計量の分布のパーセント点の高次の近似式をCornish-Fisher展開を用いて求め、非心度の漸近下側限界、漸近信頼区間も求めた。その近似式の利点はモーメントが分かれば、電卓で計算可能なもので極めて利便性が高いことである。次に、ガンマ分布、指数分布、両側指数分布、 t 分布、ワイブル分布、ロジスティック分布等で数値計算等によってその精確性について検討した。その結果、第1次近似は必ずしも良くないが、2次まで近似すれば標本の大きさが25以上で小数点以下2桁位まで合うことが分かりその精確性も確認された。また、すでに導出されている正規性の仮定の下での高次近似式との比較も行ったが、本研究の近似式は非正規性の下で考えているため少し悪くなるものの十分利用に耐えられるものであることも確認された。今後、より高次の近似式の導出が求められる。そして一般に、本研究のアプローチで、非正規性の下での様々な統計量の分布の高次近似への道が開かれたと考えられる。

さらに、臨床統計データ解析においてしばしば現れる切断分布を含む非正則分布族の場合にその母数に関して統計量を持つ情報量損失を考えることは重要でそのような場合に高次まで情報量損失を起こさないような統計量を見つけることを考えた。まず、未知の母数 θ をもつ母集団分布からの大きさ n の無作為標本に基づく統計量の(θ に関する)情報量は、元の情報量を超えないことは知られている。そしてそれらの情報量の差はその統計量の情報量損失として評価されている。通常は、分布が滑らかであるような正則条件を課した正則な場合を論じ、情報量として Fisher 情報量を用いて情報量損失について考えることが多い。しかし、正則条件が必ずしも成り立たないような非正則な分布族の場合には、Fisher 情報量は使えないので Akahira(1996)において導入された一般化情報量を用いた。ここで統計量の一般化情報量は分布の非正則性に依存する。従来、密度の台が区間であるときに、区間の両端点において密度の値は等しく、また両端点における密度の微分係数の和が0であるという条件を仮定して、極値統計量と漸近補助統計量の組から成る統計量の2次の漸近一般化情報量損失が $o(1)$ となることは示されていた。しかし、その条件は制約的なので上記の条件を課さない一般の場合に拡張することが期待されていた。本研究においてその条件を仮定しないときにも同様の結果が成り立つことを示した。このことは Akahira(1991)における統計量は2次の漸近十分であるという事実とも符合している。その拡張によってその適用範囲が大きく広がり応用への道が開かれたと考えられる。

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 12 件)

- ① Akahira, M.(2010).The first- and second-order large-deviation efficiency for an exponential family and certain curved exponential models. Commun. Statist. - Theory and Meth., 39, 1387 -1403,査読有.
- ②Akahira, M., Kim, H.G. and Ohyauchi, N.(2010). Loss of information of a statistic for a family of non-regular distributions. Mathematical Research Note 2010-001, 1-18,査読無.
- ③Kim, H. G. and Akahira, M.(2009). On the minimum variance unbiased estimation. 京都大学 数理解析研究所講究録 1621, 29-41,査読無.
- ④ 黒田晋吾・赤平昌文(2009). The approximation to the non-central

chi-square and t-distributions. 京都大学 数理解析研究所講究録 1621, 9-28,査読無.

- ⑤Akahira, M., Ohyauchi, N. and Kawai, S.(2009). A higher order approximation to a percentage point of the distribution of a non-central t-statistic without the normality assumption. Mathematical Research Note 2009-001,1-18,査読無.
- ⑥ Akahira, M. (2009). The structure of higher order asymptotic theory of statistical estimation. Amer. Math. Soc. Transl., Ser.2, 227, 175 -197,査読無.
- ⑦ 赤平昌文(2008). 非正則推定における情報量の概念とその役割. 日本統計学会誌 37(2), 329 -342,査読無.
- ⑧ 大貫光隆・赤平昌文(2007). Asymptotic properties of estimators in errors-in-variables models. 京都大学 数理解析研究所講究録 1560, 64-79,査読無.
- ⑨ 佐藤 純・赤平昌文(2007). On a comparison between the fuzzy and randomized confidence intervals. 京都大学 数理解析研究所講究録 1560, 47-63,査読無.
- ⑩ Akahira, M. and Ohyauchi, N. (2007). The asymptotic bound by Kiefer-type information inequality and its attainment. Commun. Statist. - Theory and Meth., 36(11), 2049-2059,査読有.
- ⑪ Akahira, M. and Ohyauchi, N. (2007). A Bayesian view of the Hammersley -Chapman ・ Robbins-type information inequality. Statistics.,41,137-144 ,査読有.
- ⑫Akahira, M.,Ohyauchi, N.andTakeuchi, K.(2007). On the Pitman estimator for a family of non-regular distributions. Metron 65(1), 113-127,査読有.

[学会発表] (計 6 件)

- ① 赤平昌文・Kim, H.G.・大谷内奈穂(2011). Loss of information associated with the statistic for a family of non-regular distributions. 日本数学会年会、早稲田大学、2011年3月22日
- ② 赤平昌文・大谷内奈穂・河合伸一(2010). A higher order approximation to the distribution of a non-central t-statistic without the normality assumption. 日本数学会年会、慶応義塾大学、2010年3月26日
- ③ Ohyauchi, N. and Akahira, M.(2007). Information inequalities from the viewpoint of inverse problem. The 56th Session of the International Statistical Institute, Lisboa, Portugal, Aug.27, 2007
- ④ Koike, K. and Akahira, M.(2007). Sequential estimation of a location parameter for the location-scale family of distributions in non-regular case. The 56th

Session of the International Statistical Institute, Lisboa, Portugal, Aug.27, 2007

⑤ 赤平昌文(2007). A definition of information amount and its application to non-regular cases. 日本数学会秋季総合分科会、東北大学、2007年9月24日

⑥ 大貫光隆・赤平昌文(2007). Sufficient conditions for the asymptotic normality of estimators in the errors-in-variables model. 2007年度 統計関連学会連合大会、神戸大学、2007年9月9日

[図書] (計2件)

① 広中平祐代表(2009). 現代数理科学事典 (第2版)、丸善、分担執筆「推測の漸近理論」 pp. 549-553.

② 杉山高一他編集(2007). 統計データ科学事典、朝倉書店、分担執筆「推測理論」 pp. 66-69, 74-83.

6. 研究組織

(1) 研究代表者

赤平 昌文(AKAHIRA MASAFUMI)
筑波大学・副学長
研究者番号：70017424

(2) 研究分担者

青嶋 誠(AOSHIMA MAKOTO)
筑波大学・大学院数理物質科学研究科・教授
研究者番号：90246679

小池 健一(KOIKE KEN-ICHI)
筑波大学・大学院数理物質科学研究科・准教授
研究者番号：90260471

大谷内 奈穂(OHYAUCHI NAO)
筑波大学・大学院数理物質科学研究科・助教
研究者番号：40375374

鳥越 規央(TORIGOE NORIO)
東海大学・理学部・准教授
研究者番号：40297180

白石 高章(SHIRAISHI TAKAAKI)
南山大学・経営学部・教授
研究者番号：50143160
(H19→H20：連携研究者)

田中 秀和(TANAKA HIDEKAZU)
大阪府立大学・大学院工学研究科・講師
研究者番号：50302344
(H20→H21：連携研究者)

(3) 連携研究者

今野 良彦(KONNO YOSHIHIKO)
日本女子大学・理学部・教授
研究者番号：00205577
久保川 達也(KUBOKAWA TATSUYA)

東京大学・大学院経済学研究科・教授
研究者番号：20195499