

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

研究代表者の専門分野：数理統計学

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：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. 研究計画の概要

最近、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

### 2. 研究の進捗状況

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、正規性の仮定を無くした際の非心  $t$  統計量の分布はあまり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充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。

### 3. 現在までの達成度

②おおむね順調に進展している。

(理由)

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い、順調に進んでいる。特に、医薬学分野等に応用が広い重要な2標本問題において、従来の正規性の仮定をはずした場合に、非心t統計量に基づく漸近推測の基礎を確立しつつあり、このアイデアが非正規の場合に様々な統計量の漸近挙動の解析への道を拓いたことはその証と思われる。

### 4. 今後の研究の推進方策

逐次選択実験問題において複数の処理法の適用計画まで拡張し、臨床試験へ適用可能な方向を目指す。また、2標本問題において、非正規性の下で非心統計量、標本相関係数等の漸近的性質を調べ、その応用についても検討をする。

### 5. 代表的な研究成果

(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 4 件)

①Akahira, M. (2010). The first - and second - order large - deviation efficiency for an exponential family and certain curved exponential models. In press in Commun. Statist. - Theory and Meth., 39. 査読有

②Akahira, M. (2009). The structure of higher order asymptotic theory of statistical estimation. Amer. Math. Soc. Transl., Ser.2, 227, 175 - 197. 査読有

③Akahira, 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

④Akahira, M. and Ohyauchi, N. (2007). A

Bayesian view of the Hammersley - Chapman - Robbins - type inequality. Statistics 41(2), 137-144. 査読有

[学会発表] (計 1 件)

①赤平昌文・大谷内奈穂・河合伸一: 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。

[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、xvi+1454pp.

②杉山高一他編集(2007). 統計データ科学事典、朝倉書店、xii+774pp.