

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 18 日現在

機関番号：12102

研究種目：挑戦的萌芽研究

研究期間：2012～2014

課題番号：24650146

研究課題名(和文)統計的実験とその応用の新開拓

研究課題名(英文)New Development of Statistical Experiment and Its Applications

研究代表者

赤平 昌文(Akahira, Masafumi)

筑波大学・ 名誉教授

研究者番号：70017424

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：大きさ n の標本の値域である標本空間 H を n 次元ユークリッド空間とし、これを制限して標本の最小値と最大値の凸結合の値域から成る制約標本空間を H' とし、統計的実験を標本空間、標本及びその密度の組で定義して、空間 H 、 H' に対応する実験をそれぞれ E 、 E' とする。このとき、位置母数をもつ両側切断分布の推定問題を考えると、「位置共変推定の観点からは実験 H' が H に比べて漸近的に情報量を保存する」と予想される。本研究では、この予想は否定的に解決され、一般化情報量を用いて実験 H に対する H' の漸近情報量損失を求めることができた。このことは、実験 H がもつ情報量を実験 H' では捉えきれないことを示している。

研究成果の概要(英文)：Let H be a sample space consisting of the range of sample of size n which is assumed to be the n -dimensional Euclidean space and H' a restricted sample space of the range of convex combination of the maximum value and the minimum value. Define a statistical experiment by a triple of a sample space, a random sample and its density. Let E and E' be corresponding experiments to spaces H and H' , respectively. Then we consider the estimation problem of a location parameter of a truncated distribution. From the viewpoint of equivariance, it is conjectured that there is no asymptotic loss of information associated with the experiment E' relative to E . In this research, it is negatively solved. Indeed, the value of asymptotic loss of information associated with the experiment E' relative to E was obtained. This means that the information of the experiment E can not be grasped by E' .

研究分野：数理統計学

キーワード：統計的推測 標本空間 情報量

1. 研究開始当初の背景

まず、確率変数 X が密度 $f(x, \theta)$ をもつ母集団分布に従うとすると、 X の標本空間を H とすれば母数 θ の推測のための統計的実験は $E = (X, H, \{f(\cdot, \theta)\})$ として表現される。ここで、実際に標本が得られるのが H の部分空間 H' として、制約標本空間 H' の値をとる確率変数を Y とし、その密度を $g(y, \theta)$ とすれば、 θ を推測するための Y に基づく統計的実験 $E' = (Y, H', \{g(\cdot, \theta)\})$ を考えることができる。このとき、実験 H は実験 H' より多くの情報を含んでいるように見えるが、必ずしもそうではない。たとえば、 $f(x, \theta)$ が指数型分布族の密度であるとき、一方の実験が他方の実験より Fisher 情報量が大きくなるための十分条件が求められている (Bayarri - DeGroot (1987, 1989),)。また、一定の条件の下で Fisher 情報量を最大にする母集団分布の制限方法の一つとして最適荷重関数が求められ、具体的な分布の場合について論じられている (友成・赤平 (2006))。

次に、統計量 $Y=T(X)$ の密度を $g(y, \theta)$ として、上記のように θ を推測するために Y に基づく統計的実験 $E' = (Y, H', \{g(\cdot, \theta)\})$ を考える。このとき、一般には実験 E は実験 E' より多くの情報を含むと考えられる。特に、 T として θ に対する十分統計量を取れば、 $T(X)$ は X がもつ情報量を保存する、すなわち E' は E と同じ情報量をもつという意味で同等になる。そして、 $f(x, \theta)$ が正則条件を満たすような正則な場合には、 θ の最尤推定量等を用いて高次の次数まで実験 E と同等になるような実験 E' を導く統計量を求めることができる。また、高次の次数まで漸近的に有効な推定量も求めることができる (Akahira - Takeuchi (1981), Pfanzagl - Wefelmeyer (1982))。一方、 $f(x, \theta)$ が正則条件を満たさないような非正則な場合の典型として両側切断分布族の位置母数 θ の推定問題では、最小値と最大値の組から成る極値統計量の基づく実験 E' は 1 次の次数まで漸近的に一般化情報量を保存するという意味で実験 E と同等であるが、2 次の次数では漸近的に E に対して情報損失を起こすことが示されている (Akahira (1996), Akahira et al. (2012))。

2. 研究の目的

一般に、統計学においては標本空間をユークリッド空間に取ることが多いが、実際の問題においては、標本空間を制約してその部分空間と見なせることもある。そのとき、元の標本空間上の値を取る標本に基づく実験と制約標本空間上の値を取る実験の比較や複数の制約標本空間上の値を取る実験の比較を情報量概念を通して行う。母集団分布の形状が仮定されていない一般の場合、特に非正則な場合には標本の大きさ n が無限に大きくなるような大標本論の観点から漸近情報

量損失の概念を用いて実験の比較を行う。

3. 研究の方法

統計的実験に関する先行研究の文献を詳細に調べてその関連性を含めてよく検討する。そして元の標本空間 H に対して制約標本空間 H' の構成法について考える。その際に制約の仕方が問題になり、確率変数 X の密度 $f(x, \theta)$ の形状が鍵となるが、指数型分布族のような正則分布族の範疇に入らない非正則分布族について考察する。その場合には、大標本論の観点から高次漸近理論を駆使することになるが、これまでに様々な知見を得ているので、それらを生かしていくようにする。

4. 研究成果

まず、非正則分布族の典型である位置母数をもつ両側切断分布族においてその母数 θ の推定問題を考えた。従来、このような切断分布族の場合に θ の Pitman 推定量 (最良共変推定量) の漸近展開を導出し、その漸近分散を求め、さらに切断点での密度の値による荷重推定量による比較も行われた (Akahira et al. (2007))。本研究では上記のような両側切断分布族の場合を考える。この場合には、大きさ n の標本 X の値域である標本空間 H は n 次元ユークリッド空間になり、これを制限して標本の最小値 A と最大値 B の組の極値統計量 $Y=(A, B)$ の値域と A と B の凸結合の統計量 Z の値域から成る制約標本空間をそれぞれ H', H'' とする。ここで制約標本空間 H', H'' はそれぞれ 2 次元、1 次元空間であることに注意。このとき、位置母数 θ の推定問題を考え、 θ の推定のための標本 X に基づく統計的実験 $E = (X, H, \{f(\cdot, \theta)\})$ 、統計量 Y に基づく実験 $E' = (Y, H', \{g(\cdot, \theta)\})$ 、統計量 Z に基づく実験 $E'' = (X, H, \{h(\cdot, \theta)\})$ を情報量による比較を行った。ここで情報量として非正則な場合を含む一般の場合に適用可能な Akahira (1996) で導入された一般化情報量を採用した。具体的な問題として、「位置共変推定の観点からは実験 E'' は実験 E に比べて漸近的に情報量を保存する」と予想された。本研究において、この予想を否定的に解決した。実験 E に対する実験 E'' の 1 次の漸近情報量損失を求めることができ、それは両切断点の線形結合を用いて表現された。このことは、実験 H がもつ情報量を実験 H'' では捉えきれないことを示している。一方、Akahira et al. (2012) の結果を用いれば、実験 E' は実験 E に比べて漸近情報量を 1 次の次数では保存するが、2 次の次数では漸近情報損失を起こすことが分かった。さらに、漸近補助統計量を D として統計量 $W=(Y, D)$ から成る制約標本空間を H^{\wedge} とすれば、 W に基づく実験は $E^{\wedge} = (W, H^{\wedge}, \{h(\cdot, \theta)\})$ になる。ただし、 $h(w, \theta)$ は W の密度とする。このとき、実験 E'' は実験 E に比べて漸近情報量を 2 次の次数まで保存すると

も分かった。

次に、ベイズ(Bayes)理論においては、或る事前分布について適当な推定量の値を与えた時の事後分布を求め、それを改めて事前分布として繰り返すベイズ更新が知られている。そこで、その事前分布と事後分布の差異を測ることは重要になり、その尺度としてKullback-Leibler(K-L)情報量等が知られている(Bernardo(1979), Ghosh et al.(2006))。しかし、非正則な場合にはK-L情報量はその尺度として必ずしも適用可能ではないことに注意。母数 θ の母数空間を H として、 θ の事前密度を $\pi(\theta)$ として、実験 $E=(\theta, H, \pi)$ を考える。また、 θ の最尤推定量 M を与えたときの θ の事後密度 $\pi(\theta | M)$ として実験 $E'=(\theta, H, \pi(\cdot | M))$ も考える。このとき、実験 E と E' を一般化情報量で比較した。実際に事前分布としていずれも平均0、分散1をもつ正規分布、両側指数分布、指数分布を取ったときに、実験 E と E' の間の一般化情報量を求め、それらの期待値を求めた。その結果、いずれの場合も、 E と E' の間の情報量の差異は比較的小さいことが分かった。また、特に一様事前分布が一般一様分布に近づけば、実験 E と E' の間の情報量の近似差が無くなることも分かった。

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 7 件)

¹ 大谷内奈穂・赤平昌文(2014). Second order asymptotic behaviour of location equivariant estimators for a family of truncated distributions. 京都大学 数理解析研究所講究録, 査読無, 1910, 102-114, <http://www.kurims.kyoto-u.ac.jp/~kyoto/kokyuroku>

² 赤平昌文・小池健一(2014). The generalized amount of information between the prior distribution and the asymptotic posterior one. 京都大学 数理解析研究所講究録, 査読無, 1910, 144-153, <http://www.kurims.kyoto-u.ac.jp/~kyoto/kokyuroku>

³ Akahira, M.(2014). Interval estimation procedures and information inequalities. Proc. 59th ISI World Statistics Congress, 査読無, 1120 - 1125, <http://www.statistics.gov.hk/wsc/STS002-P1-A.pdf>

⁴ 大谷内奈穂・赤平昌文(2013). Asymptotic comparison of estimators for a family of truncated distributions. 京都大学 数理解析研究所講究録, 査読無, 1860, 129-139, <http://www.kurims.kyoto-u.ac.jp/~kyoto/kokyuroku>

⁵ 橋本真太郎・小池健一(2013). ベイズリスクに対するバッチャリヤ型不等式. 京都大学 数理解析研究所講究録, 査読無, 1910, 1-8, <http://www.kurims.kyoto-u.ac.jp/~kyoto/kokyuroku>

⁶ 矢田和善・青嶋 誠(2013). On the distribution of the largest eigenvalue in high dimension, low sample size context. 京都大学 数理解析研究所講究録, 査読無, 1860, 120-128, <http://www.kurims.kyoto-u.ac.jp/~kyoto/kokyuroku>

⁷ 赤平昌文(2013). 歴史的現象の統計的考察と画期的な推測法の進展. 日本物理学会誌, 査読有, 68, 81-88.

[学会発表](計 6 件)

¹ 大谷内奈穂・赤平昌文: Asymptotic comparison of location equivariant estimators for a family of truncated distributions. 2014年度統計関連学会連合大会, 2014年9月15日, 東京大学本郷キャンパス(東京都文京区)

² 赤平昌文・大谷内奈穂: Asymptotic loss of the MLE of a truncation parameter for a one-sided truncated exponential family of distributions. 日本数学会年会, 2015年3月23日, 明治大学駿河台キャンパス(東京都千代田区)

³ 大谷内奈穂・赤平昌文: Asymptotic loss of the MLE of a truncation parameter for a two-sided exponential family of distributions. 日本数学会年会, 2015年3月23日, 明治大学駿河台キャンパス(東京都千代田区)

⁴ 橋本真太郎・小池健一: Reference priors with maximum divergence for multiparameter non-regular models. 日本数学会年会, 2015年3月23日, 明治大学駿河台キャンパス(東京都千代田区)

⁵ 矢田和善・青嶋 誠: 高次元データにおける拡張クロスデータ行列法の漸近的性質とその応用について. 日本数学会年会, 2015年3月24日, 明治大学駿河台キャンパス(東京都千代田区)

⁵ Akahira, M.: Interval estimation procedures and information inequalities. The 59th World Congress of the International Statistical Institute, Aug. 29, 2013, Hong Kong Convention Exhibition Centre, China

⁶ 矢田和善・青嶋 誠: PCA consistency for the power spiked model in high-dimensional settings. 第7回日本統計学会春季集会, 2013年3月3日, 学習院大学(東京都豊島区)

[その他]

ホームページ:

<http://www.math.tsukuba.ac.jp/~akahira/masafumi/>

6. 研究組織

(1) 研究代表者

赤平 昌文 (AKAHIRA MASAFUMI)
筑波大学・名誉教授
研究者番号：70017424

(2) 研究分担者

青嶋 誠 (AOSHIMA MAKOTO)
筑波大学・数理物質系・教授
研究者番号：90246679

小池健一 (KOIKE KEN-ICHI)
筑波大学・数理物質系・准教授
研究者番号：90260471

大谷内奈穂 (OHYAUCHI NAO)
筑波大学・数理物質系・助教
研究者番号：40375374