様　式　C－１９、F－１９－１、Z－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ３０ 年　 ６ 月 １３ 日現在

機関番号：１３６０１

研究種目：基盤研究(C)（一般）

研究期間：2013～2017

課題番号：２５３８０２６５

研究課題名（和文）多変量分布が作るリーマン空間の幾何学的考察

研究課題名（英文）Riemannian manifold of multivariate distributions in view of geometrical aspects

研究代表者

椎名　洋（Shiina, Yo）

信州大学・学術研究院社会科学系・教授

研究者番号：８０２４２７０９

交付決定額（研究期間全体）：（直接経費）　　2,500,000円

研究成果の概要（和文）：１）一般的なパラメトリック分布の母数を最尤法で推定し、その推定量を挿入して得られる分布（予測分布）と真の分布の乖離をアルファーダイバージェンスで計測し、その期待値の漸近展開を求めた。一次の項は、母数の数/2であり、二次の項は、多様体の幾何的な性質（計量、接続、曲率など）によって決定される。多項分布、正規分布（分散・共分散未知）、混合分布について、具体的な二次の項を求めた。２）重回帰分析において、誤差項がパラメトリックな分布に従う時に、上の一般論を適用して、推定の効率を考察した。特に、漸近展開の二次の項に現れる幾何的な性質が、説明変数の分布や誤差項の分布によってどう変わるかを考察した。

研究成果の概要（英文）：1) Consider a family of parametric distributions. We get a predictive distribution by inserting the maximum likelihood estimator. We measure the discrepancy the predictive distribution and the true distribution by alpha-divergence and define the risk as its expectation. We derived an asymptotic expansion w.r.t. the sample size. The first term ,i.e. 1/n term equals (the number of the parameters)/2 and the second term, i.e. 1/n term is expressed with various geometrical properties. We gained concrete results for a multinomial distribution, a multivariate normal distribution, a mixture family. 2) We applied the above general result to a multiple regression model with a parametric error distribution. We investigated how the distribution of the explanatory variables and the error distribution affect the geometrical properties in the second term.

研究分野：数理統計学

キーワード：情報幾何　リーマン多様体　ダイバージェンス　曲率　接続

１．研究開始当初の背景

（１）Amari[1,2], Amari & Nagaoka[3], Murray & Rice[4]に端的にまとめられているように、パラメーターによって規定される確率分布のなす族は、そのパラメーターを座標とし、Fisher 情報統計量を計量とすることで、リーマン空間になる。さらに、そこに「アルファー接続」を導入することで、平行移動、曲率等の幾何学的諸量も賦与される。こうした幾何学的諸量が推定論や検定論における高次漸近理論と深く結びついているという結果（Amari & Kumon [5], Kumon & Amari[6]）は、情報幾何の重要な成果の一つである。

しかしながら、本研究計画策定時（平成24年頃）において、多変量正規分布、多変量ｔ分布等の具体的な分布において、こうした諸量が具体的にどのような関数になるか、またその関数が統計的推測において示唆するものの意味はまだ十分に解明されていなかった。特に、漸近論をはなれた小標本における正確な分布論や、指数分布族の枠組みにあてはまらない分布において、それらの諸量が果たす役割はほとんど解明されていなかった。

（２）当時、研究代表者は、平均０の多変量正規分布の作るリーマン空間（分散共分散行列＝正値対称行列をパラメーターに持つ空間）をスペクトル分解の視点から、その幾何学的構造を探った自分自身の研究（Sheena[7]）で、次のような成果を得ていた。１）分散共分散行列の固有値・固有根を新しい座標とした時のリーマン計量、２）固有値、あるいは固有ベクトルのどちらかを固定して得られる部分多様体の曲率、３）固有値を推測するにあたって、標本固有値のみを使用することから生まれる情報量損失、４）幾何学的視点から見た自然な固有値推定量。これらの結果は、従来の古典的な分析方法では得られない、新しい結果であり、上記の未解明な課題の一部に対して答えを与えたものであった。本研究は、この Sheena[7]で行われたアプローチを拡張することをねらったものであった。

＜引用文献＞

[1] S. Amari. “Differential geometry of curved exponential families-curvature and information loss.” The Annals of Statistics, 10: 357-385,1982.
[2] S. Amari. “Differential-Geometrical Methods in Statistics.” Lecture Notes in Statistics 28. Springer-Verlag, 1985.
[3] S. Amari and H. Nagaoka. “Methods of Information Geometry”. Translations of Mathematical Monographs 191. American Mathematical Society, 2000.
[4] M. K. Murray and J. W. Rice. “Differential Geometry and Statistics.” Taylor & Francis, 1993.
[5] S. Amari and M. Kumon. “Differential geometry of Edgeworth expansions in curved exponential family.” Annals of the Institute of Statistical Mathematics, 35: 1-24, 1983.
[6] M. Kumon and S. Amari. “Geometrical theory of higher-order asymptotics of test, interval estimator and conditional inference.” Proceedings of the Royal Society of London, A387: 429-458, 1983.
[7] Y. Sheena. “Inference on the eigenvalues of the covariance matrix of a multivariate normal distribution -- geometrical view --“, Journal of Statistical Planning and Inference, 150: 66-83, 2014.

２．研究の目的

（１）当初の具体的な目的は、以下のとおりであった。

ア）他の多変量分布に、Sheena[7]でとったアプローチをあてはめることで、正規分布と同様の成果を得る。

イ）指数分布族でない分布、特に多変量ｔ分布のような分布で幾何学的な諸量と推測がどう結びついているかを探る。

ウ）ある分布族全体の曲率（そのなかに埋め込まれた curved exponential family のような部分多様体の曲率ではなく）がそもそも統計的推測にとって何を意味するかを考察する。

（２）しかし、本研究のテーマは、「情報幾何の枠組みのなかで、多変量分布の統計的推測に関して、何らかの新たな知見を得る」ことであり、このテーマに沿った何らかの理論的成果が得られる可能性がある場合は、そちらを優先することも視野にいれていた。

３．研究の方法

理論的な研究が中心なので、

ア）既存の知識・結果を論文・参考書から学ぶこと、

イ）それらの応用または拡張・発展を考えること、

この二つを交互に行っていくことが中心となった。また、何らかの理論的な結果が得られた場合に、

ウ）シミュレーションによって結果を確認することも、度々行った。

本研究は、「研究代表者」単独での研究であったが、学会・研究会等の場において、関連した研究をフォローすること、他の研究者とディスカッションすることも行った。

４．研究成果

（１）「研究の目的」の（ウ）について、次のような知見を得た。この成果は、「主な発表論文等」の〔雑誌論文〕の②の論文として公刊された。また、〔学会発表〕の①、②において、発表された。

様々な統計モデル（パラメトリックモデルに限る）を相互に比較して、どのモデルが推定しやすいのか、また同じモデルにおいて、

パラメーターのどのあたりの値において、推定は困難になるのか。この二つに関して、パラメーターの最尤推定量をパラメーターに代入した時の予測分布と真の分布のずれをアルファーダイバージェンスで計測し、その期待値で推定の評価を行うといったアプローチをとり、次のような成果を得た。
①評価式の漸近展開を標本数の逆数の二次の項まで求めた。一次の項は、(パラメーターの数/標本数)/2 となる。二次の項については、モデルをリーマン多様体として考えた場合の計量・接続・曲率等の幾何的諸量で表現される。これによって、漸近的にではあるが、任意のパラメトリックモデル相互の、あるいはモデル内での「推定のしやすさ」の比較が可能になった。
②いくつかの素朴なモデル(多項分布、多変量正規分布、混合分布)に関して、上記の二次の項を計算し、どのような場合に推定がしやすいか(難しいか)を考察した。また、二項分布と比較して、そのモデルの推定がどの程度困難かを測るというアイデアも提唱した。
(2)「研究の目的」の(ア)と(イ)について、次の様な知見が得られた。この成果は、「主な発表論文等」の〔雑誌論文〕の①の論文として公刊された。また、〔学会発表〕の②において、発表された。(1)で得られた一般論を、回帰モデルというより具体的なモデルに応用して、より具体的な形で「推定のやりやすさ」を計測したのがこの研究成果である。
　重回帰分析は、統計的手法の中で、最もよく使用されているものの一つであるが、誤差項が、正規分布以外のものである時に、漸近的な意味での推定効率がどうなるかについて、具体的な結果はまだ多く得られていなかった。これに関して、次のような成果を得た。
①重回帰分析において、誤差項の分布を既知(位置母数と尺度母数は不明)とした場合に、最尤推定量で係数と分散を推定する。この結果から得られる予測分布と真の分布のずれをアルファ―ダイバージェンスで計測し、その期待値を推定の評価に使う。この評価式を標本数の逆数で展開した場合の二次の項までを算出した。(1)で述べた通り、一次の項の係数は、常にパラメーターの数(回帰モデルの場合、係数の数+2)の半分であるが、二次の項は、誤差項の分布と説明変数の分布(モーメント)に依存している。具体的な誤差項として、ア)正規分布、イ)t 分布、ウ)歪正規分布の三つ、一方で説明変数の分布として、ア)正規分布、イ)t 分布、ウ)二値分布、エ)パレート分布の三つ、これらの組み合わせ計 12 通りについて、二次の項を計算した。
②得られた評価式と二項分布(B(n,p))に関する同様の評価式を比較することで、標本数に関する二つのメルクマールを提案した。一つは、回帰と二項分布で(パラメータ数/標本数)が等しい状況で、二つの評価式が等しくなるような二項分布のｐを求め、これがどの程度 0 に近いかを見るもの(I.D.E.)。もう一つは、二項分布で p=0.5,n=10 とした時に、二つの評価式が一致する回帰側の標本数を見るもの(R.S.S.)である。
③結論として、回帰分析は総じてコイン投げ(二項分布)と同等あるいは少し高い程度の推定のしやすさをもつモデルであることが分かった。但し、誤差項や説明変数のモーメントの選び方によって、推定の困難さはかなり異なり、さらにアルファーにも依存する。

(3)今回の研究において、アルファーダイバージェンスによる分布間の距離を測ることは基本的なアプローチであるが、それに関して、「研究目的」の(ア)~(ウ)に該当しないが、情報幾何の観点から重要な問題について次のような知見を得た。
　ある測度に従属する実数値確率分布を、有限次元の多項分布で近似することは、応用上大変重要である。その際、区間を固定して、その区間とそれに対応する確率からなる多項分布として近似するか(固定区間の方法)、分位点を用いて、分位点からなる区間とそれに対応する確率からなる多項分布として近似するか(浮動区間の方法)は、二つの対照的な方法である。この二つの漸近的な推定効率(アルファーダイバージェンスで近似分布と本当の多項分布のずれを測り、その期待値をとったものの漸近展開)を比較し、次の結果を得た。
①標本数の逆数の一次の項の係数は、固定区間、浮動区間いずれの場合も、区間の数/2 となる。
②アルファーダイバージェンスとマイナスアルファーダイバージェンスの平均をとったものを、分布のずれの計測に使うと、浮動区間による近似が漸近的に固定区間による近似をドミネイトする(二次の項が常に小さい)。

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３件)

① Yo Sheena, “Asymptotic expansion of the risk of maximum likelihood estimator with respect to α-divergence”, ***Communications in Statistics - Theory and Methods-***,Vol.47,4059-4087,2018,査読有

②Yo Sheena, “Asymptotic expansion of risk for a regression model with respect to alpha-divergence with an application to the simple size problem”, ***Far East Journal of Theoretical Statistics***, Vol. 53,

Number 4, 187-230, 2017, 査読有

③ Yo Sheena, " Inference on the eigenvalues of the covariance matrix of a multivariate normal distribution -- geometrical view --" , ***Journal of Statistical Planning and Inference***, Vo.150, 66-83, 2014、査読有.

〔学会発表〕（計２件）

①椎名　洋、"Asymptotic Expansion of Divergence between Parametric Distributions"、統計数理研究所セミナー、2016

②椎名　洋、「パラメトリックな確率分布間の Divergence の漸近展開」、2015 年度統計関連学会連合大会、2015

６．研究組織

(1)研究代表者

椎名　洋（Shiina, Yo）
信州大学・学術研究院社会科学系・教授
研究者番号：80242709