氏　　　　名	髙橋勇人
学位（専攻分野）	博士(学術)
学位記番号	総研大甲第569号
学位授与の日付	平成１４年3月２２日
学位授与の要件	数物科学研究科　統計科学専攻 学位規則第４条第１項該当
学位論文題目	パラメトリックモデルとコルモゴロフ複雑度
論文審査委員	主　査　教授　田邉　國士 教授　伊藤　栄明 助教授　栗木　哲 助教授　土谷　隆 教授　伊藤　俊次（津田塾大学） 教授　合原　一幸（東京大学） 教授　釜江　哲朗（大阪市立大学）

論文内容の要旨

本論文は，0，1の無限列上の確率測度とKolmogorov complexityとの関係を探究したものである。特にパラメトリックモデルから自然に得られる計算不可能な確率測度とKolmogorov complexityとの関係について論じている。

計算可能な確率測度とKolmogorov complexityとの関係は,これまで知られているが，計算不可能な確率測度とKolmogorov complexityとの関係はあまり知られていない。本論文ではこれを調べるために，A.R.Barron [Ph.D.Dissertation 1985] によって示されたBayes符号を，Kolmogorov complexityの評価に適用する方法を選んでいる。それは，パラメトリックモデル$P(\cdot;\theta)$が与えられたとき，直接Shannon符号を用いて符号長を評価するのではなく，$P(x_n) = \int P(x_n;\theta)d\theta$のようにパラメータで積分した測度にShannon符号を適用するものである。この符号化方法をKolmogorov complexityの評価に適用するには，$\bar{P}$が計算可能であると仮定すればよい。“パラメータが計算可能 ⇔ モデルが計算可能”という状況のもとで，この仮定はかなり一般的に成立することがら知られている。またこのときの符号長の上限はモデルが滑らかであるという条件の下ですべてのパラメータに対して

$$-\log P(x_n;\hat{\theta}) + \frac{k}{2}\log n + o(\log n) \tag{1}$$

となることが知られている。ただし，$\hat{\theta}$は最尤推定量を表しkはパラメータの次元とする。また符号長の下限もルベーグ測度0のパラメータの集合を除いてほぼ同じオーダとなることが知られている。

本論文では，上記のBarronで示された方法を次ぎの2点にわたって拡張している。
1　モデルが滑らかという条件の下でパラメータがルベーグ測度0の集合に属する場合の符号長，およびKolmogorov complexityの評価をしたこと。

2　モデルが滑らかでない場合のKolmogorov complexityの評価をしたこと。特に無理数回転モデルから得られる確率モデルを探究し，そのKolmogorov complexityの評価をするとともに，このモデルが滑らかでないことを示し，またそれがもつ幾つかの特徴を明らかにした。

上記の1を調べるために特異測度を事前分布に採用している。積分論においては一般に特異測度による積分の理論が展開されてはいるが，具体的に特異測度による積分の計算をした例はあまりないので，この点に本論文の貢献がある。また，尤度関数が最尤推定量のまわりに集中するという統計学のモデルでは標準的な条件のもとで特異速度による積分の漸近展開を計算している。これによってこれまでに知られていた事実を自然な形で一般化した結果が得られ，滑らかなパラメトリックモデルの符号長の構造がかなり解明されている。特に本論文では，具体的に計算できる特異測度とKolmogorov complexityから導かれる有限な集合体上の測度を用いたBayes符号の符号長の漸近展開を計算している。その結果，パラメータがルベーグ測度0の集合に属する場合,（1）の$\log n$オーダの項が通常のBayes符号と比べて小さくできることを示している。

上記の2に関しては，力学系から導かれるパラメトリックモデルの単純で興味深い例となっている。特に，無理数回転の中でも興味深い性質をもつSturmian列を生成するモデルの性質が考究されている。その結果，このモデルが通常の滑らかなパラメトリックモデルと比べて著しく異なる特徴をもつことが判明した。例えば，尤度関数のサポートが縮退すること，尤度関数のグラフの形が全て3角形となることなどが証明されている。またこのモデルから生成される列，すなわちSturmian列のKolmogorov complexityの評価が得られ，滑らかなモデルとは異なることが示されている。またパラメータが計算不可能な実数の場合，生成されるSturmian列は全て計算不可能であることを明らかにしている。

論文審査結果の要旨

本論文は、コルモゴロフ複雑度と計算不可能な確率測度の間の関係について論じたものである。

コルモゴロフ複雑度は0-1列に関する複雑度をチューリングマシンでその列を出力するのに必要な最小なプログラムの長さとして定めたものである。0-1列に対する確率測度 $P(x)$ を考える。任意の計算可能な $P(x)$ が与えられた時、文字列 x に対するコルモゴロフ複雑度と $-\log P(x)$ の間には漸近的な関係があることが知られており、シャノン符号や MDL 符号、ベイズ符号等に基づいてコルモゴロフ複雑度の上界と下界を評価することができる。ただし、確率測度が計算可能であるとは、任意の与えられた文字列に対して、任意の精度でその確率を計算するアルゴリズムが存在するということを意味する。これまでのコルモゴロフ複雑度と確率測度の関係の研究では確率測度が計算可能であるということを前提としていた。これに対し、本論文では、パラメトリックモデルにおいて、確率測度がパラメータの関数として滑らかな場合について、パラメータの事前分布が特異であるようなベイズ符号を考え、その符号長を評価している。一般に、適当な条件の下で、パラメトリックモデルの符号長の下限を達成する符号は MDL 符号等や (事前分布に適当な正則性を仮定した通常の意味での) ベイズ符号で与えられる。ここで提案され、解析されているベイズ符号は、以下のように構成される特異な事前分布に基づくものである。まず、パラメータ θ が "それを2進展開した時に、0と1の出る確率が $p \in (0,1)$ であるような実数の集合上に分布が集中しているような特異事前分布" を考え、"それをさらに p に関して0と1の間で積分して得られるような事前分布" を考える。そして、この事前分布とルベーグ測度とを1対1の比率で混合したものを事前分布として採用する。この符号の符号長を、最尤推定量を用いて評価すると、MDL 符号においてパラメータに関する情報を記述するために現れる罰金項 $(k/2)\log n$ (k: パラメータの次元、n: サンプル数) に相当する項が、MDL 符号よりも小さくなることが示される。特に、p が0あるいは1に近づく時に、(これはパラメータが有限桁で表現できる場合に相当する) 符号長はシャノン符号の符号長に近づき、1/2に近づく時に MDL 符号の符号長に近づく。この意味で、提案されている符号はシャノン符号と MDL 符号とをつなぐものであると見ることもできる。さらに、この結果を一般化することも試みられている。本章で得られている結果は符号長やコルモゴロフ複雑度の研究に新しい見方を与えるものとして興味深いものである。

本論文ではさらに、確率測度がパラメータの関数として滑らかでない場合として、無理数回転に関連したスツルム列と呼ばれる0-1列について論じ、そのコルモゴロフ複雑度を評価している。長さ1の円周を考え、これを区間 $[0,1)$ と自然に同一視する。適当な初期点から出発して、一定長 ρ だけ円周上を右周りに動くことを繰り返して生成される点列 $\{x^{(k)}\}$ を考える。この時、$x^{(k)} \in [1-\rho, 1)$ であれば1、そうでなければ0として点列 $\{y^{(k)}\}$ を生成する。この点列 $\{y^{(k)}\}$ を $(x^{(0)}, \rho)$ から生成されるスツルム列と呼ぶ。スツルム列は、周期性を有しない一番単純な0-1列として知られ、古くからそれについて多くの研究がなされてきた。申請者は、0-1列を生成する確率モデルとして ρ を固定した時に、初期値 $x^{(0)}$ が一様分布に従うとして得られるスツルム列を考え、生成される点列のコルモゴロフ複雑度を評価することを試み、その上下限などについていくつかの興味深い結果を得た。コルモゴロフ複雑度の上下限を評価するには尤度関数の性質を調べることが必要になる。申請者はこのモデルの尤度関数を2次元平面上に巧妙に図示する方法を考案し、それを用いて凸性などいくつかの興味深い性質を導いた。さらに、考案した図に基づき、尤度関数のサポート部分は三角形である

という著しい性質があることを予想した。この予想は、釜江哲朗氏との共同研究において実際に証明された。この図自身は、すでに発見されていたことが後に判明したが、尤度関数のサポート部分が三角形であるという性質は興味深い新しい結果であり、申請者が現在さらに研究を進めているスツルム列に関する統計的推論の理論の基礎になるものである。以上述べてきた問題意識やアプローチはスツルム列研究の一つの新しい方向を示唆したものとして、高く評価される。また、この展開が既存の研究対象を確率モデルや尤度などの観点から見直すことで初めて可能になったことは、数理科学における統計科学的な視点の重要性を示している。

以上のように、申請者は、自ら研究テーマを設定し、独自の視点を持ってじっくりと腰を据えて研究を続け、上述のような興味深い研究成果を得た。審査委員会は申請者の研究者としての能力を高く評価し、学位を授与することが適切であると判断した。

なお本論文の内容に関して、申請者は、2つの論文をすでに投稿しており、その内1つはすでに掲載が決定している。また、もう1つは査読者の報告に基づき現在改訂中である。さらに、もう1本の論文を現在投稿準備中である。