

早稲田大学大学院 基幹理工学研究科

博士論文概要

Theory of Multivariate Time Series Analysis and Its Application

多変量時系列に対する理論とその応用

申請者

Hideaki	NAGAHATA
長幡	英明

数学応用数理専攻 数理統計・時系列・金融研究

2018年12月

本学位論文では、多変量時系列および高次元時系列に対する判別解析や分散分析といった手法における古典的統計量について、望ましい性質を得るための正則条件を明らかにし、判別の一致性や検定統計量の漸近分布を導出する。また、多変量時系列および高次元時系列を解析する新たな判別統計量や検定統計量も提案し、その漸近的性質について議論している。さらに、これらの新たな理論的結果の応用として、日本国内における上場企業の財務データ・対数株価データ、そして福島県の環境放射線量データに対するそれらの応用例も紹介している。

本学位論文の構成は、以下の通りである：

Chapter 1. 導入

Chapter 2. 多変量および高次元時系列における判別解析とクラスター解析

Chapter 3. 高次元金融時系列の分類

Chapter 4. 多変量時系列に対する分散分析

Chapter 5. 高次元時系列に対する分散分析

Chapter 6. 高次元時系列分散分析モデルにおける検定統計量の高次の近似分布

以下、各章の概要を述べる。

Chapter 1 では、本学位論文における研究の動機と具体的研究対象である「判別解析・クラスター解析」や「分散分析」といった統計的手法の歴史および潜在する問題点を指摘し、Chapter 2 から Chapter 6 の導入を次のように述べている。多変量解析において、多変量データを分析するための様々な方法論が提案・開発されており、これらの古典的結果は理論発展上も実用上も大変重要である。しかし、それらの多くはデータ同士の独立性を仮定した場合の結果であるため、「多変量時系列」および変量数が無限である「高次元時系列」において古典的結果を適用することはできない。多変量時系列および高次元時系列に対する理論的結果およびその応用について議論する。具体的には、歴史が長く・頻繁に用いられる手法である「判別解析・クラスター解析」および「分散分析」に焦点を当て、新たな理論を構築するとともに、金融データや放射線データに対するそれらの応用例を紹介する。

Chapter 2 では、「判別解析・クラスター解析」に関して、特に多変量および高次元時系列データに対する判別解析を扱った。具体的には、平均 μ 、スペクトル密度 $f(\lambda)$ 、自己共分散行列 $R(s) = \{R_{ij}(s); i, j = 1, \dots, p\}$ の定常過程 $\mathbf{X} = \mathbf{X}(t) = (X_1(t), \dots, X_p(t))'$ を考える。この定常過程の観測系列 $\{\mathbf{X}(1), \dots, \mathbf{X}(n)\}$ が次の 2 つの群 π_1 、 π_2 ：

$$\pi_1: \mu = \mu^{(1)}, \quad f(\lambda) = f^{(1)}(\lambda), \quad R(t) = R^{(1)}(t),$$

$$\pi_2: \mu = \mu^{(2)}, \quad f(\lambda) = f^{(2)}(\lambda), \quad R(t) = R^{(2)}(t),$$

のいずれかに属するとする。また、構造を記述する π_1 、 π_2 も未知であるが、 π_1 からこのような 2 群が異なる母平均をもつ非正規な定常過程と仮定し、標本数・次元数が共に大きくなる高次元の場合を考え、高次元独立標本で応用された判別統

計量について議論した。結果としては、高次元従属標本において前述した判別統計量の誤判別確率 $\rightarrow 0$ なる条件を解明した。また、この判別統計量が高次元従属過程では無視できないバイアスをもつため、従属標本に対する修正判別統計量を提案した。以降、判別統計量のバイアス中に現れる自己共分散行列を **Jackknife** 法によって推定した、この修正判別統計量を「**Jackknife** 型修正判別統計量」と呼ぶ。**Jackknife** 法とは標本から重複をしないように・標本のいくつかを抜いてリサンプリングを繰り返し、得られた多数の標本の集まりから母集団の性質（今回は自己共分散）を推定する手法である。こちらの結果としても、高次元従属標本において「**Jackknife** 型修正判別統計量」の誤判別確率 $\rightarrow 0$ なる条件を解明した。さらに、「上場企業らの財務情報」を元に本手法に基づく「樹形図」を作り、信用格付けへの貢献を見据えた応用例も示した。

Chapter 3 では、Liu, Nagahata, Uchiyama, Taniguchi (2018)で提案された手法を応用し、高次元金融時系列の分類を行っている。近年、Aoshima and Yata (2014) といった高次元独立標本に対する判別解析は研究されているものの高次元従属標本に適用できる解析手法は Liu, Nagahata, Uchiyama, Taniguchi (2018)で提案された「**Jackknife** 型修正分類器」などをはじめとし、いまだ発展途上の段階である。ましてや、そういった高次元従属標本の現実問題への理論適用に関してはまったく未成熟な状況となっている。昨今ビッグデータが話題になっているように実際の金融データでも高次元時系列データはますます増えており、このような分析は強く要請されている。本章では、様々な業界から代表的な上場企業 24 社（一部上場企業 12 社・二部上場企業 12 社）を選定し、公開されている貸借対照表・損益計算書・キャッシュフロー計算書といった財務情報 15 年分を高次元時系列データとみなし、これらの 24 社のクラスター解析を実施した。「古典的分類器」と「**Jackknife** 修正型分類器」のクラスター解析結果を比較し、一部上場企業の中でも規模の大きい企業に対して「**Jackknife** 型修正分類器」がより適切な分類結果となっていることを論じている。

Chapter 4 では、一元配置の多変量分散分析モデル：

$$\mathbf{X}_{it} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{it}, \quad i = 1, \dots, q, \quad t = 1, \dots, n_i$$

を扱う。 $\{\mathbf{X}_{it}, t = 1, \dots, n_i\}$ を*i*群における長さ n_i の観測系列とし、 $\boldsymbol{\mu}$ はモデルの平均、 $\boldsymbol{\alpha}_i$ を*i*群の処理効果、 $\{\boldsymbol{\epsilon}_{it}, t = 1, \dots, n_i\}$ を*i*群の攪乱項とする。従来とは異なり攪乱項は従属過程と想定して古典的検定統計量の帰無仮説下での漸近分布（カイ二乗分布）を導出した。従属過程については正規性を仮定せず、非正規定常過程として議論をする。定常過程とは時間通じて一定の平均まわりで変動しておりトレンドをもたない過程のことである。研究結果としては、まず従属標本の下で前述した古典的な検定統計量が漸近的にカイ二乗分布にしたがい、その際の十分条件は攪乱項が無相関過程であることを示した。これは攪乱項の想定する従属過程が相関をもつ場合に古典的な検定が使用できないことを意味するため、この十分条件なしで漸近的にカイ二乗分布にしたがう **Whittle** 尤度に基づく検定統計量を新しく提案した。**Whittle** 尤度は定常過程が正規性をもつと仮定して構成した尤度の近似である。また、このような漸近理論だけでなく、現実のデータに適用する際

に必要となる「有限標本での検定統計量の性能」をシミュレーションによって確認した。帰無仮説下 ($\alpha_i = \mathbf{0}$) で、 ϵ_{it} を無相関過程や (無相関過程はない) 従属過程として発生させた観測系列に対し棄却率が有意水準と近くなる標本数を数値的に求めた。さらに、対立仮説下 ($\alpha_i \neq \mathbf{0}$) でのシミュレーションにより有限標本下での検出力も評価した。これらの理論および数値実験結果により、正当性を確認した上で、実際の金融データを用いた分析も行った。アメリカの証券取引市場に上場されているいくつかの企業の日次対数株価に対して「その業界平均が等値性をもつか」を検定するため、古典的な検定と Whittle 尤度に基づく新しい検定を使って分散分析をした。検定結果は大きな差異が出なかったものの、日次対数株価データの標本自己相関関数を調べると強い相関をもったデータだと判明したため、Whittle 尤度型の検定こそが今回のようなデータに応用できることを示唆している。

Chapter 5 では、高次元従属標本の場合に古典的統計量の漸近分布やその十分条件を明らかにすることを目的とする。近年、観測機器などの発達に伴い高次元データ、ここではデータの次元数 (変量) が大きいデータへの解析需要が高まっている。Fujikoshi et al. (2011) は高次元独立標本に対して分散分析における古典的統計量の漸近分布が種々の仮定の下でカイ二乗分布となることを導出した。しかしながら、独立標本に対して開発された手法が高次元従属標本に適用できるかどうかは自明ではなく、未解明かつ重要な問題である。よって、本章では、Chapter 4 と同様に一元配置の多変量分散分析モデル ($\{\mathbf{X}_{it}, t = 1, \dots, n_i\}$ を i 群における長さ n_i の観測系列とし、 μ はモデルの平均、 α_i を i 群の処理効果、 $\{\epsilon_{it}, t = 1, \dots, n_i\}$ を i 群の攪乱項) を扱うものの、攪乱項は「高次元従属過程」を想定している。高次元従属過程については正規性を仮定せず、Brillinger (1975) と同様にキムラントに関する自然な条件 (以降、Brillinger 条件と呼ぶ) を付した非正規定常過程として議論をし、古典的検定統計量の帰無仮説下での漸近正規性を導出した。同時に、古典的検定統計量は無相関過程である場合、漸近正規性を有することを解明した。これにより、従来の分散分析が解析対象とできなかった時系列データを解析できる。実データに対する分散分析の例として、多時点で観測された福島県の放射線量時系列データを分析した。

Chapter 6 では、高次元従属過程の攪乱項をもつ一元配置の多変量分散分析モデルにおいて、Chapter 5 で導出した古典的検定統計量の漸近分布よりも精度の高い漸近分布を導出することを目的とする。前章での古典的検定統計量の正規近似は一次近似であったが、本章ではさらに高次の近似を導出することで精度を向上させる。具体的には、攪乱項に Brillinger 条件などの適切な正則条件を課した上で、Edgeworth 展開を用いることによって、古典的検定統計量の帰無仮説下での漸近分布を三次近似まで導出した。さらに、次のようなシミュレーションによって、有限標本でも Edgeworth 展開を用いた漸近分布が前章の正規近似よりも優れていることを確認した。帰無仮説下 ($\alpha_i = \mathbf{0}$) で、 ϵ_{it} を無相関過程として発生させた観測系列から検定統計量の経験分布を得た後、その経験分布を基準として「Edgeworth 展開により得た新たな近似分布」と「正規近似」とを比較したところ、今回導出した「新たな近似分布」がより小さい誤差をもつことを示した。

早稲田大学 博士（工学） 学位申請 研究業績書

氏名 長幡 英明 印

(2019年 1月 29日 現在)

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
単著論文	○Nagahata, H. (-). Higher order approximation of the distribution of test statistics for high-dimensional time series ANOVA models. To appear in <i>Scientiae Mathematicae Japonicae</i> .
共著論文	○Nagahata, H. and Taniguchi, M. (2018). Analysis of variance for high-dimensional time series. <i>Statistical Inference for Stochastic Processes</i> , 21(2):455--468.
共著論文	○Nagahata, H. and Taniguchi, M. (2018). Analysis of variance for multivariate time series. <i>METRON</i> , 76(1):69--82.
単著論文	○Nagahata, H. (2017). Classification for high-dimensional financial time series by a class of disparities. <i>Advances in Science Technology and Environmentology Special Issue</i> , 2017. Vol.B14, 47-55.
共著論文	○Liu, Y., Nagahata, H., Uchiyama, H., and Taniguchi, M. (2017). Discriminant and cluster analysis of possibly high-dimensional time series data by a class of disparities. <i>Communications in Statistics-Simulation and Computation</i> , 46(10):8014-8027.
共著論文	○Nagahata, H., Suzuki, T., Usami, Y., Yokoyama, A., Ito, J., Hasegawa, F., Taniguchi, M. (2012). Faces for financial time series data. <i>Advances in Science Technology and Environmentology Special Issue</i> , 2012, B8:1-13.