

## 科学研究費助成事業 研究成果報告書

平成 27 年 6 月 4 日現在

機関番号：12601

研究種目：挑戦的萌芽研究

研究期間：2013～2014

課題番号：25610030

研究課題名(和文)大規模不等間隔時空間データの推測理論

研究課題名(英文)Statistical inference theory for irregularly spaced spatio-temporal data

研究代表者

矢島 美寛(YAJIMA, Yoshihiro)

東京大学・経済学研究科(研究院)・教授

研究者番号：70134814

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): 定常時空間確率場の不等間隔地点(時点)で観測されるデータに基づいて統計的仮説検定を行う場合に対して、新たな検定統計量を提案し、その理論的性質特に大標本における漸近的性質について研究した。この統計量の長所は、帰無仮説の下で極限分布が正規分布にしたがうと仮定した場合、その期待値と分散がスペクトル密度関数の推定に用いるカーネル関数を用いて、簡潔な表現になることである。したがって現実の大規模データに応用する際にも、時間節約的な検定方法として貢献できる。極限分布が正規分布になることの証明は現在鋭意努力中である。

研究成果の概要(英文): We proposed a new test statistic for statistical hypothesis testing given irregularly spaced spatio-temporal data from a stationary random field.

If we assume that the limiting distribution under the null hypothesis is a normal distribution, its expectation and variance are expressed by a much simple form in terms of the kernel function that is used to construct an estimator of the spectral density function. Hence this statistic is applied easily to analyze an actual large spatio-temporal data. We are trying now to justify theoretically that the limiting distribution is normal.

研究分野：統計科学

キーワード：時空間統計解析 時空間定常確率場 統計的仮説検定 不等間隔時空間データ

### 1. 研究開始当初の背景

時空間統計データは環境学、疫学から都市経済学に至るまで広範な分野において存在する。たとえば環境学においては地球温暖化問題に関連して、地球各地点における時間的な温度変化、疫学においては伝染性疾患の感染状況の時空間的变化、都市経済学においてはある都市の産業集積の状況あるいはある地域の地価家屋の価格が近隣都市や近隣地域との時空間的相互作用によりどのように変遷するかなど解明すべき問題が多々ある。

それらを解析するための時空間モデルも様々なものが提案されている。しかし通常の独立同一分布を仮定したモデル、時系列解析における定常性を仮定したモデルに比べ、そのモデルの複雑さから、当てはめられたモデルの妥当性、統計学的に言えば検定についての研究は数少ない。

### 2. 研究の目的

(1) 上述の背景の下で、時空間モデルに対して新たな検定統計量を提案する。

(2) 提案した統計量の理論的性質、特に大標本理論における漸近的性質を明らかにする。

(3) 理論的性質を導出する際には、以下の2点に留意する。

日次、週次など等間隔時点で観測される場合が多い時系列データと異なり、時空間データを観測する際には、観測地点・時点が不等間隔な場合が多い。またそれらは固定されておらずランダムに選択されると仮定する方が現実的な場合がある。このような観測状況における理論展開を試みる。

リモートセンシングにより観測されるデータなどのように、時空間データの個数は膨大な数になることがある。このような大規模データに対しても、迅速に時間節約的に計算可能な検定統計量を提案する。

### 3. 研究の方法

(1) 時空間モデルとして、今までに様々なモデルが提案されている。その代表的な統計モデルで実際のデータ解析にも頻繁に応用されるモデルの一つとして時空間定常確率場がある。時空間定常確率場とは、各時点・各地点における観測値の期待値が一定で、異なった地点間・時点間における観測値間の共分散関数が地点・時点をベクトルで表現したとき、そのベクトル差のみの関数で決定されるモデルである。

本研究では時空間定常確率場に対するモデルの適合度を検定する新たな検定統計量を提案する。

もちろん現実のデータには期待値が一定ではない、あるいは共分散がベクトル差のみでは決まらない非定常性を示すデータは多々存在する。そのようなデータに対しては、期待値の部分は各地点・各時点の特徴を表す変数の関数で表現し、誤差項に定常確率場を仮定した回帰モデルや空間的あるいは時間的に区域を分割し、各区域における観測値には局所的に定常性を仮定したモデルが応用

されることも多い。したがって時空間定常確率場は、他の時空間モデルの出発点に位置するモデルであり、本研究ではこのモデルに絞って研究を遂行した。

その場合当てはめたモデルの妥当性すなわち統計的仮説検定について議論した研究はきわめて少ない。研究開始当初の背景で述べたようにこのような状況の下に、従来の統計学の基本的枠組みすなわち独立同一分布あるいは定常過程にしたがいかつ等間隔時点で観測されるデータに対して発展してきた統計的推測理論を直接適用することはできない。さらに観測地点・時点が固定されておらずこれもランダムな確率変数になる場合、理論的展開は格段に難しくなる。たとえば原データは多変量正規分布にしたがう正規定常確率場であっても、観測地点・時点がランダムになるともはや正規定常確率場にはならない。このような状況の下で不等間隔地点(時点)において観測されかつ大規模なデータに対しても時間節約的な検定統計量を構築し、その理論的性質特に大標本における漸近的性質を明らかにすることが目的である。(2) 時空間定常確率場の自己共分散関数(自分自身の異なった地点・時点における観測値間の共分散関数なのでこのように呼ばれる)はスペクトル密度関数と呼ばれる関数の多変数フーリエ変換によって表現できる。したがって当てはめたモデルの適合度を判定するには、そのモデルのもとでのスペクトル密度関数と真のスペクトル密度関数の乖離の大きさを測れば良い。

(3) 真のスペクトル密度関数は未知であるから推定する必要がある。通常時系列解析では、まずデータが等間隔で得られる場合には様々な周波数においてデータをフーリエ変換(離散フーリエ変換と呼ばれる)し、ピリオドグラムと呼ばれるその絶対値の2乗を計算する。真のスペクトル密度関数の推定量として、ピリオドグラムはサンプルが多くなるにつれてバイアスは0に収束し、漸近的に不偏な推定量ではあるが、分散は0に収束せず一致推定量ではない。そこで分散が0に収束するように、着目している周波数に対するピリオドグラムと、その近傍の周波数に対するピリオドグラムをカーネル関数によって平滑化した推定量を採用する。この推定量はカーネル関数にある仮定を置くと、サンプル数が多くなるにつれてバイアス・分散ともに0へ収束する一致推定量であることが証明されている。この推定量を時空間定常確率場におけるスペクトル密度関数の推定にも応用する。

(4) しかしながら前述のような観測地点(時点)が不等間隔なおかつランダムな場合には上述の時系列解析の結果が成立するか否かは自明ではない。このような観測状況に置けるピリオドグラムおよび平滑化されたスペクトル密度関数の理論的性質をまず明らかにする必要がある。

(5)ピリオドグラムおよび平滑化されたスペクトル密度関数の理論的性質を明らかにした後、新たな検定統計量を構築する。ここでは推定されたスペクトルの密度関数とその帰無仮説の下での期待値の差の2乗を様々な周波数において計算し、ある周波数の領域において積分する。積分する際には尺度不変にするために、帰無仮説の下でのスペクトル密度関数で上述の2乗を割っておく。そしてこの積分が大きい場合には帰無仮説を棄却する。

この検定統計量の理論的性質を明らかにすることが本研究の目的である。

#### 4. 研究成果

(1)ピリオドグラムは周波数間の距離が一定以上であるならば、異なった周波数におけるピリオドグラムは時空間定常確率場であっても、さらに観測地点(時点)が不等間隔かつランダムであっても、定常過程から等間隔時点で観測された時系列データによって構築されたピリオドグラムと同様に、漸近的には、無相関であることを明らかにした。

(2)したがってカーネル関数を用いて平滑化したスペクトル密度関数の推定量も、時系列解析の場合と同様に一致推定量になることを明らかにした。

(3)スペクトル密度関数の推定量とその帰無仮説の下での差の2乗に基づく検定統計量に対して、帰無仮説の下で極限分布が正規分布に従うと仮定した場合、その期待値と分散を導出した。観測地点(時点)はある有界な領域において、独立同一分布に従う確率変数列から選択されかつ観測値の個数とともに領域が増大していくと(mixed asymptotics と呼ぶ)、上述の期待値と分散は未知の量に依存せず、カーネル関数の多重積分および観測地点(時点)を決定する確率分布関数から決まる定数のみによって表現できる。

(4)したがって大規模なデータに対しても、迅速に時間節約的に計算することが可能である。

(5)対立仮説の下では、検定統計量はサンプル数が増加すると無限大に発散し、したがって一貫性を持つ。

(6)なお検定統計量としてはスペクトル密度関数の推定量とその帰無仮説の下での差の2乗に基づくものだけではなく、テーラー展開したとき2次の項が主要項となるような関数たとえばKullback-Leibler 情報量、J ダイバージェンス、Chernoff インフォーマションなどに基づく検定統計量についても同様の理論的結果が成立することを明らかにした。

(7)現在は、下記スペインで開催されたコンファレンスにおいて Professor Suhashini Rao(Texas A&M University)からいただいたコメントに基づき、極限分布が正規分布にしたがうことの理論的正当化を示すのに鋭意努力中である。これが証明できれば、理論的

にも応用上も、国内外においてかなりのインパクトを与えると期待できる。

(8)最後に今後の展望・課題について列挙する。

時系列解析の場合にも同様の状況が生じるが、観測値の個数が増加しても検定統計量の分布がその極限分布に収束するスピードが遅い場合がある。その場合代替的な計算機集約的な手法としてブートストラップ法がある。時系列解析においては、等間隔で観測されたデータからのスペクトル密度関数の推定量に基づくブートストラップ法の理論的妥当性はすでに証明されている。一方時空間定常確率場に関しては、データそのものに基づく推定量に対するブートストラップ法についてはその理論的妥当性を証明した論文はあるが、スペクトル密度関数の推定量に基づく場合についてはまだその理論的妥当性は証明されていない。この点を明らかにするのが今後の一つの課題である。

次に複数の検定統計量の優劣を明らかにする一つの方法として、局所対立仮説の下での検出力を比較する方法がある。上述の複数の検定統計量の優劣をこの方法を用いて明らかにする。

次に3.研究方法(5)で述べた積分する周波数領域は、観測値の個数が増大しても固定されている。観測値の個数とともに領域も増大させていく方が、さらに優れた検定統計量を構築できると予想できるが、この点を理論的に明らかにする

研究者代表者はかつて時空間定常確率場の実際データへの応用として、公示地価データに対して、私鉄を利用した場合の最寄り駅から東京都区内の主要 JR 駅までの時間距離、最寄り駅から自宅までの地理的な距離などを説明変数とし、誤差項に定常確率場を仮定した回帰モデルを適用して分析した。誤差項の自己共分散関数には方向には依存せず距離のみに依存する関数(等方型自己共分散関数という)を帰無仮説に採用して、尤度比検定を行った。本研究で提案した検定統計量の結果との異同を調べることで、また両検定統計量の優劣を理論的に明らかにすることも今後の課題である。

#### 5. 主な発表論文(研究代表者、研究分担及び連携研究者には下線)

[雑誌論文](計 1 件)

Yoshihiro Yajima, On statistical testing for spatio-temporal random fields, Abstract of the 2nd international conference of the International Society of Nonparametric Statistics, (査読なし), 1, 2014, 55-56.

[学会発表](計 2 件)

Yoshihiro Yajima, On statistical testing for spatio-temporal random

fields. The 3rd Institute of Mathematical Statistics Asia Pacific Rim Meeting, Taipei, Taiwan, 2014.06.30-2014.07.03.

Yoshihiro Yajima. On statistical testing for spatio-temporal random fields. The 2nd international conference of the International Society of Nonparametric Statistics (invited speaker) Cadiz, Spain, 2014.06.12-2014.06.16

## 6 . 研究組織

### (1)研究代表者

矢島美寛 (YAJIMA Yoshihiro)  
東京大学・大学院経済学研究科・教授  
研究者番号：70134814

-