

# 時系列解析におけるロバスト統計計算

RA 清 智也

情報理工学系研究科 数理情報学専攻 博士課程 2 年

## 概要

本研究では、ある種の時系列モデルが持っている局所漸近混合正規性 (LAMN) と呼ばれる性質に着目し、この性質を持つモデルのための新しい情報量規準を導出した。導出した規準が従来のものに比べ、リスクの意味でロバスト性を持つことを数値的に確認した (1 節)。また、LAMN モデルの一例である、拡散過程の離散観測モデルにおいて、この規準の安定な計算法を示した (2 節)。さらに、拡散過程の離散観測モデルは、空間データのためのモデルへ拡張可能であることを示した (3 節)。

## 1 新しい情報量規準

近年、金融工学やロボット工学、ニューラルネットワークなどを含む多くの分野において、拡散過程モデルに基づいた統計的解析が盛んに行われている。ここで、拡散過程とはランダムネスを含む連続時間の力学系のことであり、確率微分方程式によって記述される。このような系の離散的な観測値に基づき確率微分方程式の拡散項を推定するという問題も、応用上重要な研究対象の 1 つとなっている。拡散項の関数形は通常未知であり、色々なモデルのあてはめが試みられる。その中から最適なモデルを選ぶためのモデル選択規準 (= 情報量規準) は、データに隠れている本質的な構造を探る上で非常に重要な役割を果たしている。

モデル選択規準として最も広く用いられているものに赤池情報量規準 (AIC) がある。この AIC を導出する際に前提となっている仮定は、局所漸近正規性 (LAN) である。LAN は、標本数の増加とともに正規分布モデルで近似できることを意味

し、独立同一分布モデル、回帰モデル、定常時系列モデルなど多くのモデルが兼ね備えている性質の 1 つである (表 1)。しかし、先に挙げた拡散過程の離散観測モデルは一般に LAN ではなく、代わりに局所漸近混合正規性 (LAMN) を有する。LAMN は、標本数の増加とともに混合正規分布モデルで近似できることを意味する。

表 1: LAN と LAMN

漸近的性質	LAN	LAMN
極限モデル	正規分布	混合正規分布
例	独立同一分布, 回帰, 定常時系列	非定常 AR, 分枝過程, 拡散過程

このような背景から、本研究ではまず、任意の LAMN モデルに対し適用できる形の情報量規準 LAMN-Bayes-IC を導出した ([1], [2])。ここで、Bayes という名前がついているのは以下の理由による。情報量規準は、一般に予測分布のリスクの推定量として定義される。ここで予測分布として何をを用いるかが問題となる。ベイズ予測分布 (= ベイズ事後分布に基づく予測分布) は、プラグイン予測分布 (= モデルの未知パラメータに最尤推定量をプラグインして得られる予測分布) に比べ一般に良い性能を持つことが知られており、その差は LAMN モデルでは特に顕著である。従ってプラグイン予測を用いることは得策でなく、ベイズ予測を用いることにした。

数値実験の一例を図 1 に示す。ここでは LAMN モデルの極限である混合正規分布モデルに対して、

LAMN-Bayes-IC と AIC を用いたときのリスクを数値的に計算した。横軸は真のパラメータを、縦軸はモデル選択におけるリスクを表している。詳しい説明は省略する。この結果、LAMN-Bayes-IC は AIC に対し、比較的リスクが小さいことが確かめられる。図中の BOUND とは、真のパラメータが既知という条件下で最適なモデルを選択したときのリスクである。LAMN-Bayes-IC は、BOUND との差が比較的一定であるという点でロバストであると言える。拡散過程の離散観測モデルに対しても、同様のシミュレーション結果が得られた。

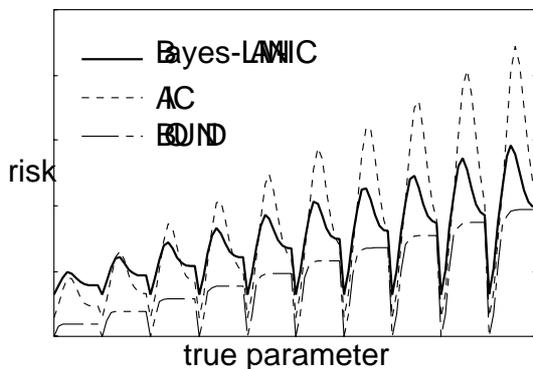


図 1: 数値実験の結果.

## 2 ロバストな計算法

情報量規準 LAMN-Bayes-IC の計算において、モンテカルロ法に基づく期待値の計算が必要となる。つまり、AIC のように負の最大対数尤度とパラメータ数を加えるとといった比較的簡単な計算ではなく、繰り返し計算が必要となる。この期待値の計算の際、数値的安定性の観点から問題となるのは、観測情報行列 (= 対数尤度のヘシアン) の逆行列を計算する部分である。観測情報行列は一般に正定値行列であるが、最尤推定量の近似などにより正定値でない行列が現れる場合がある。特に、LAN モデルと異なり、LAMN モデルの場合は観測情報行列が確率変数に収束するため、特異に近い行列が現れる確率は少なからず存在する。

この問題に対し、モデルを拡散過程モデルの場合に限れば、以下のような対処が可能である。すな

わち、観測情報行列の極限はある非負定値行列の積分形で表される、という性質に着目し、この積分を差分近似した量を計算するという対処法である。これにより、比較的ロバストな LAMN-Bayes-IC の計算手法を獲得することができた。

## 3 空間統計モデルへの拡張

拡散過程の離散観測モデルの拡張として、ブラウン布 (Brownian sheet) を変換して得られる空間統計モデルを考えることができる。ただし、変換関数には未知パラメータが含まれるものとする。このモデルは LAMN であることが、本研究により示された。結果は現在まとめている段階である。LAMN 性が成り立つので、LAMN-Bayes-IC に基づくモデル選択が可能となる。

このモデルは、比較的研究されていない非ガウス型の空間統計モデルの一例になっている。変換関数には制限がないことから、実際のデータ解析においても有用であると考えられる。また、ブラウン布以外のガウス過程を変換する場合についても現在考察を進めており、幅広い応用を見込んでいる。

## 4 おわりに

1 節, 2 節の結果については、2003 年 12 月に香港において行われた国際会議 [1] で発表し、好評を得た。また、3 節で述べた内容を含め、現在論文を執筆・投稿中である。

## 参考文献

- [1] Sei, T. and Komaki, F. (2003). An information criterion for models with the local asymptotic mixed normality, Proceedings of Bernoulli Society East Asian and Pacific Regional Conference, December 18-20, 2003, HKUST, Hong Kong, p. 101.
- [2] Sei, T. and Komaki, F. (2004). Bayesian prediction and model selection for locally asymptotically mixed normal models, In preparation.