

情報量規準を援用した時系列データの平滑化

池本博美 (指導教員: 吉田裕亮)

1 はじめに

私たちの身の回りには気象情報, 経済現象の記録, 医学データといった様々な時系列データが存在する. これら時系列データの解析方法として, 一般的には時系列モデルに含まれるパラメータの推定が行われている. また, 時系列で平均, 分散, 共分散が時間的に変化しないものを定常時系列といい, 定常でないデータは非定常時系列と呼ばれる.

通常, データ解析では定常時系列を扱うことが好ましいものの, 現実には扱う時系列データは非定常性である場合が多くそのままではモデル解析を行うことが難しい. そこで本研究では, 情報量規準を援用し非定常時系列を定常時系列に平滑化する手法を提案する.

2 フーリエ変換

$f(x) \in L^1(\mathbb{R})$ に対して, 以下の式をフーリエ変換という.

$$\hat{f}(\xi) = \int f(x) \exp(-2\pi i x \xi) dx$$

フーリエ変換は, 時間を変数とする関数を周波数を変数とする関数に変換する. その応用領域は, 信号処理, データ処理, 通信など多岐にわたる. 通常, フーリエ変換で扱われるのは標準化されたデジタル信号であり, 有限回の計算で終わらせる必要がある. このような制約を受けたフーリエ変換を, 特に離散フーリエ変換という.

長さ n のリスト u_r の離散フーリエ変換 v_s は

$$\frac{1}{\sqrt{n}} \sum_{r=1}^n u_r \exp \frac{2\pi i (r-1)(s-1)}{n}$$

で定義される. 逆に, 周波数関数から時間関数に変換する手法を逆離散フーリエ変換といい, 長さ n のリスト v_s に対する逆離散フーリエ変換 u_r は

$$\frac{1}{\sqrt{n}} \sum_{s=1}^n v_s \exp \frac{-2\pi i (r-1)(s-1)}{n}$$

で定義される. 本研究ではこれらを用いる.

3 赤池情報量規準 AIC

モデル解析を行う際は, 最尤法を用いる場合が多い. 最尤法とは, 対数尤度を最大とする θ を選ぶことでモデルのパラメータを推定する手法である. しかし, 異なるモデル間の比較においては推定量に偏りが生じてしまうため, 最大対数尤度そのままでは利用できない.

そこで, モデル選択のための評価基準として用いられるのが AIC(Akaike Information Criterion) であり, 以下の式で表される.

$$\begin{aligned} AIC &= -2 \log L(\theta) + 2k \\ &= -2(\text{最大対数尤度}) + 2(\text{パラメータ数}) \end{aligned}$$

AIC を最小にするモデルを選択することにより, 最適なモデルが得られる. 本研究では, フーリエ変換による平滑化の最適な規準を調べるために AIC を援用した以下の規準を用いることとする.

$$n \log \sigma^2 + 7d$$

ここで n はデータ数, σ^2 は各モデルの分散, d は残周波数の個数とする.

4 提案手法

本研究では, 定常時系列に変換するための平滑化の閾値を, 実データを用いて次のように検証する.

1. 東京 / 大阪 / 網走 / 那覇の各地点における 2000 年 1 月 1 日から 2016 年 12 月 31 日までの 17 年分の日平均気温データを, 離散フーリエ変換を用いて周波数関数に変換しパワースペクトルを求める.
2. ある閾値で周波数を残して平滑化し, 周期成分と定常成分に分解する.
3. 残った定常成分の分散と残周波数の個数に提案規準を当てはめ, 値が最小になる場合を最適解とする.
4. 他の各地点においても同様の実験を行い, 全体として最適な平滑化の基準を見つける.

5 実験結果

(I) 東京

図 1 は, 東京の 17 年分の気温データである.

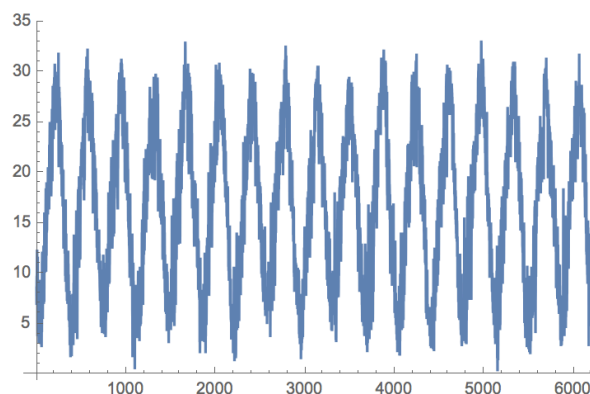


図 1: 東京の気温データ

図 2 は, この実データからフーリエ変換を用いて求めたパワースペクトルの図となっており, 右がその拡大図である.

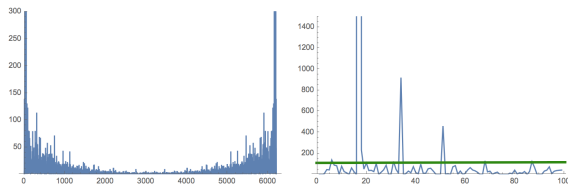


図 2: 左: パワースペクトル / 右: 拡大図

平滑化の閾値	分散	周波数の数	提案規準
120	3.502	16	7895.84
130	3.514	12	7889.24
135	3.520	10	7885.62
140	3.526	8	7882.27
400	3.536	6	7885.59

表のように、様々な値で平滑化を行ったところ、周波数を 8 個残して平滑化した場合に提案規準が最小となることがわかった。

図 3 は実データの一部を拡大した図と、実際に閾値 140、残周波数 8 で平滑化した図である。

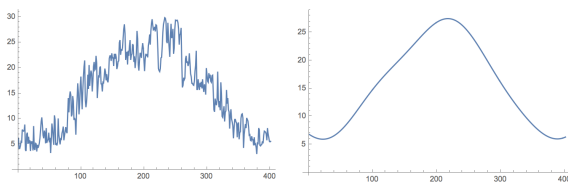


図 3: 左: 気温データ / 右: 周波数 8 で平滑化後

ここで、提案基準で定めたパラメータが妥当であるかを検証するため、17 年間の東京の日平均気温データから各日における 1 年分の平均気温を算出し、同様の手順で平滑化を行なった。

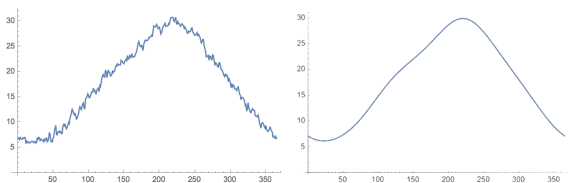


図 4: 左: 平均気温 / 右: 周波数 8 で平滑化後

また、この平均気温データに対して 15 日ごとの移動平均をとったグラフは図 5 の左のようになる。図 5 右は、周波数を 8 個残して平滑化した図 4 右と重ね合わせたものであり、この場合が当てはまりが最も良く、最適な平滑化であると判断できた。

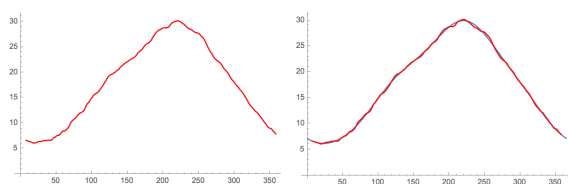


図 5: 左: 移動平均 / 右: 重ね合わせ

大阪, 網走, 那覇についても同様の手順で検証した。

(II) 大阪

平滑化の閾値	分散	周波数の数	提案規準
100	2.493	14	5771.37
120	2.497	12	5769.07
160	2.503	10	5768.67
180	2.509	8	5770.23
200	2.516	6	5771.93

(III) 網走

平滑化の閾値	分散	周波数の数	提案規準
260	5.719	12	10913.75
300	5.731	10	10912.28
330	5.743	8	10911.99
380	5.757	6	10913.15
500	5.775	4	10918.07

(IV) 那覇

平滑化の閾値	分散	周波数の数	提案規準
75	1.407	14	2220.84
80	1.410	12	2219.80
85	1.413	10	2219.52
103	1.417	8	2221.76
105	1.420	6	2223.29

表より、大阪は周波数 10 個、網走は周波数 8 個、那覇は周波数 10 個を残して平滑化するとそれぞれ提案規準が最小となることがわかった。

6 まとめ

本研究では、非定常な実データに対し、情報量規準を援用して平滑化の閾値を決定することで定常時系列に変換した。今回使用した気温データにおいては、閾値を 8 から 10 の間に設定することで最も適切に平滑化できることが確認できた。本研究は、時系列データのモデル解析を行う前処理として有効な手法であると考えられる。今後はアウトライヤーも考慮した成分分解を行うことで、より精度の高い解析を目指したい。

参考文献

- [1] 北川源四郎著、「時系列解析プログラミング」
- [2] 北川源四郎, 石黒真木夫, 坂元慶行共著, 「情報量統計学」
- [3] 気象庁, 過去の気象データ・ダウンロード, <http://www.data.jma.go.jp/gmd/risk/obsdl/index.php>
- [4] 若浦雅嗣, 「一般化加法モデルを用いた気温過程の平均・分散構造解析」, <http://www.applstat.gr.jp/jjas/33-2-4.pdf>