研究紹介

蛭川　潤一

新潟大学・理学部・数学科

時系列

時と共に変動する偶然量の観測値の系列を時系列という

数学的にはこの系列を
１つの確率過程（確率変数の族）の実現したものとみなす

確率過程の統計解析を時系列解析という

通常の統計学の議論は主に独立標本に対する議論である

独立標本 $X_1, X_2, \ldots, X_n, \ldots \Rightarrow$ 独立同一分布（i.i.d.）

時系列

時系列解析は過去、現在、未来の系列が
互いに従属している（影響しあっている）状況
での統計解析である

$\Rightarrow$ より一般的な設定のもとでの統計解析の議論である

$X_1, X_2, \ldots, X_t, \ldots$ 従属（dependent）

$\Downarrow$

確率過程 $\{X_t\} \Longrightarrow X_1, \ldots, X_n$ 観測系列が得られる

$\uparrow$ $\swarrow$

構造についての意見を述べる

実際の時系列データ

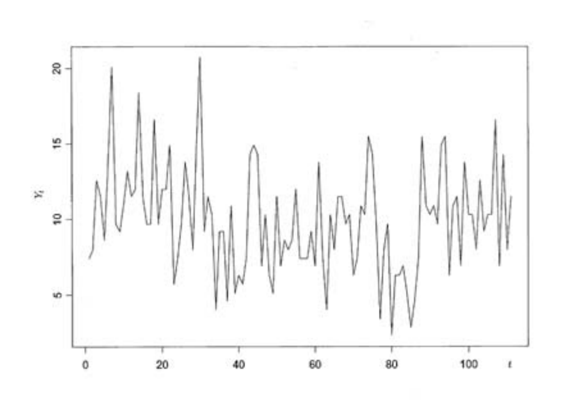

図 1 : ニューヨークでの連続 111 日間の風速（マイル/時）データ
$Y_1, Y_2, \cdots, Y_{111}$

実際の時系列データ

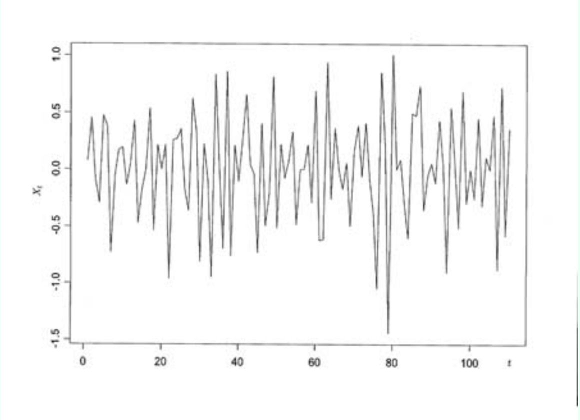

図 2：Y_t の対数差分 $X_t \equiv \log Y_{t+1} - \log Y_t,\ t = 1, 2, \cdots, 110$

標本自己相関関数

観測系列： $X_1, X_2, \cdots, X_n$ が得られたとき

標本自己相関関数 sample autocorrelation function

$$SACF(l) \equiv \frac{\sum_{t=1}^{n-l}(X_{t+l} - \bar{X}_n)(X_t - \bar{X}_n)}{\sum_{t=1}^{n}(X_t - \bar{X}_n)^2}$$

の動きをみることが多い
ただし、 $\bar{X}_n = n^{-1}\sum_{t=1}^{n} X_t$

X_{t+l} と X_t の相関の強さを表す指標

$\{X_t\}$ が互いに独立、あるいは無相関であれば
$\Rightarrow SACF(l) \approx 0$ for $l \neq 0$

標本自己相関関数

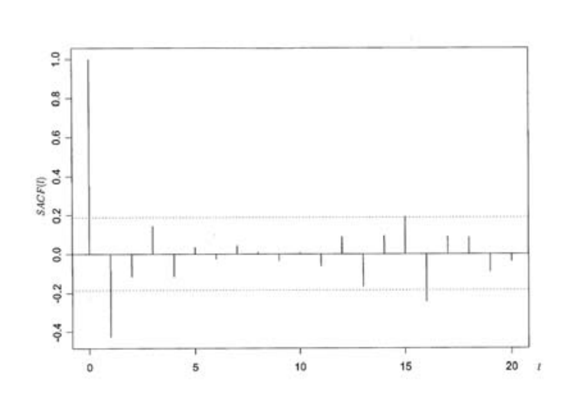

図 3：図 2のデータ $X_1, X_2, \cdots, X_{110}$ の $SACF(l)$, $l = 0, 1, \cdots, 20$

標本自己相関関数

$SACF(l)$ は $l \neq 0$ のときもかなり大きい値をとる l がある

$\Downarrow$

$\{X_t\}$ が互いに独立、あるいは無相関であるとは想定し難い

$\Downarrow$

このようなデータに対して
どのような時系列モデルを構成すればよいであろうか？

自己回帰過程

$\Downarrow$ 回帰分析的な考えに立てば

X_t がそれ自身の過去の値 $X_{t-1}, \cdots, X_{t-p}$ の線形結合と誤差項 u_t の和で表されるモデル

(1) $$X_t = -b_1 X_{t-1} - \cdots - b_p X_{t-p} + u_t$$

を思いつくだろう
ここに、 $\{u_t\} \sim i.i.d\ (0, \sigma^2)$

$\{X_t\}$ は p 次の自己回帰過程 pth order autoregressive model (AR(p)) と呼ばれる

以後、$\{X_t\} \sim \mathrm{AR}(p)$ と表記する

自己回帰過程

$\Rightarrow$ 従属データへのモデルとして、最も簡単で説明力のあるモデル

上述の風速データ $X_1, X_2, \cdots, X_{110}$ が
(1) 型のモデルに従っていると想定する

次数 p 係数 $b_1, \cdots, b_p$ と u_t の分散 σ^2 は未知だから
データから推測しなくてはならない

自己回帰過程

$\Downarrow$ （詳細を省いて結果だけ述べると）

標準的な推測法を用いて
未知パラメータ $(p, b_1, \cdots, b_p, \sigma^2)$ を推定すると

その推定値は

$$(\hat{p}, \hat{b}_1, \cdots, \hat{b}_p, \hat{\sigma}^2) = (4, 0.6452, 0.5079, 0.2233, 0.1766, 0.1647)$$

となり、$\{X_t\} \sim \mathrm{AR}(4)$ となった

現時点の（気象現象の）値 X_t が
過去 4 時点前までの値 $X_{t-1}, \cdots, X_{t-4}$ に
影響を受けていることを意味している