

第2章 時間領域における単変量解析 .....	1
2.1. 変動の持続性 .....	1
2.2. 時系列モデリング .....	2
2.2.1. 赤色ノイズ .....	2
2.2.2. 赤色ノイズの生成方法 .....	4
2.2.3. 自己回帰過程(autoregressive process) .....	5
2.2.4. 自己回帰モデルの推定 .....	6
2.2.5. 移動平均(MA)過程 .....	8
2.2.6. 自己回帰移動平均(ARMA)過程 .....	8
2.2.7. 自己回帰和分移動平均(ARIMA)過程 .....	8
2.2.8. コレログラム .....	8
2.3. 統計推定・検定の自由度 .....	8
2.3.1. 相関に関する無相関時間の導出 .....	11
2.4. フィルタリング .....	13
参考文献 .....	14

## 第2章 時間領域における単変量解析

### 2.1. 変動の持続性

普通の統計の教科書では、 $N$  個の**独立な**データがあつてなんとかは自由度  $N-1$  のかんとか分布に従う、と何気なく書いてある。「独立な」は書いてない場合も多い。しかし、7月が猛暑であれば多分8月も暑いと予想できるように、気温など大気・海洋の統計データ解析で扱うデータは、時間・空間的に独立ではなくである程度の**持続性(persistence)**を持つ。その持続時間よりも短いサンプリング間隔であれば、データを独立とみなすことができず、近くのデータ同士は遠くのデータ同士よりも強く相関する。この性質を、**連続相関(serial correlation)**があると言う。連続相関という用語はただし、具体的な相関係数を意味するのではなく、連続相関があるということの実際の計算はラグ相関などで行なわれている。

持続性をもたらすメカニズムの一つは海洋の熱的な慣性である。すなわち、2m の厚さの熱容量が大気全層の熱容量に匹敵する海洋の温度が変化すると、その記憶は長く保たれる。海洋表面水温の経年変動が数ヶ月に渡って持続することは、Namias ら (1988)が研究し、この論文は 1970 年代の気候レジームシフトについての先駆け研究としても有名である。しかも、熱的慣性はしかし数ヶ月後に失われるわけではない。大気からの熱的強制の偏差によって、冬季の海洋混合層内に温度偏差が作られると、浅い夏季の混合層の下で温度偏差が保存され、また次の年の冬季に表面水温に影響を与える。このメカニズムを、SST 偏差の再発現 (re-emergence)と呼ばれ、Alexander ら(1999)などによって研究されている。このような熱的慣性が存在すれば、それが大気にフィードバックすることで、大気もまたある程度の過去の状況に引きずられると考えられる。また、海洋の熱的な慣性によらなくても、大気自身のブロッキングも過去の記憶を未来に残すメカニズムとなっている。

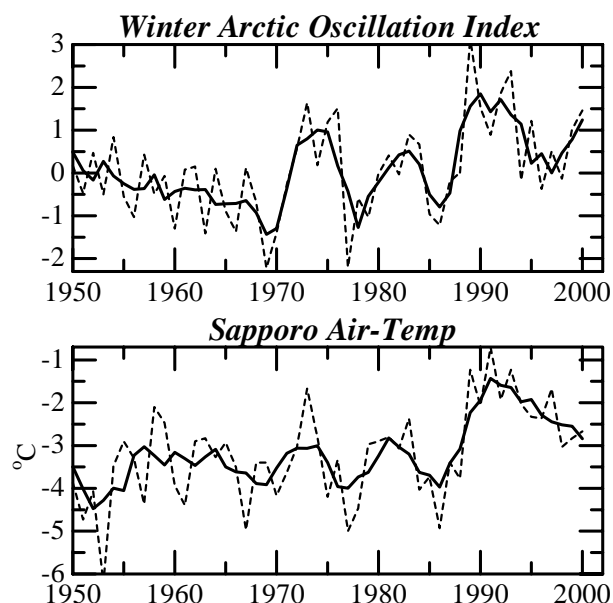


図 2.1 冬季の北極振動指数（上）と，札幌の気温（下）．点線が各年のデータで，実線が5年移動平均．各年が独立とは言い難い．

## 2.2. 時系列モデリング

### 2.2.1. 赤色ノイズ

持続性を持つ現象について，何事かをモンテカルロ法によって調べようとするれば，Matlab の `randn` で表現される乱数は，適切な時系列データの代理時系列とはなり得ない．この場合に最も広く使われかつ最も簡単な時系列の生成モデルは，次式で示される**赤色ノイズモデル(red noise model)**である．

$$y(t) = ry(t-1) + \sigma_w \varepsilon(t) \tag{2.1}$$

ここで時系列  $y(t)$  を表現する．ここで， $t$  は離散化された時間ステップ， $r$  はラグ1自己相関係数， $\varepsilon$  は分散1のガウスノイズ． $\sigma_w$  はガウスノイズの標準偏差である．後に示すように，赤色ノイズは長周期成分のスペクトルが強いために，光のスペクトルとの類推から赤色ノイズと呼ばれる．またガウスノイズは，**白色ノイズ(white noise)**とも呼ばれる．なお上の式で， $r=0$  とすれば，白色ノイズに他ならないので，赤色ノイズは白色ノイズを含んでいる．また，容易に見て取れるように， $r$  が大きいほど持続性も大きい．

赤色ノイズの実現例を図 2-1. に示す．一見周期的に見える長期変動が赤色ノイズには存在し得るが，後に示すように赤色ノイズは特定の周期性を持たない（スペクトルピークが無い）ので，視察では周期性があるように見るとしても偶々そう見えるだけである．

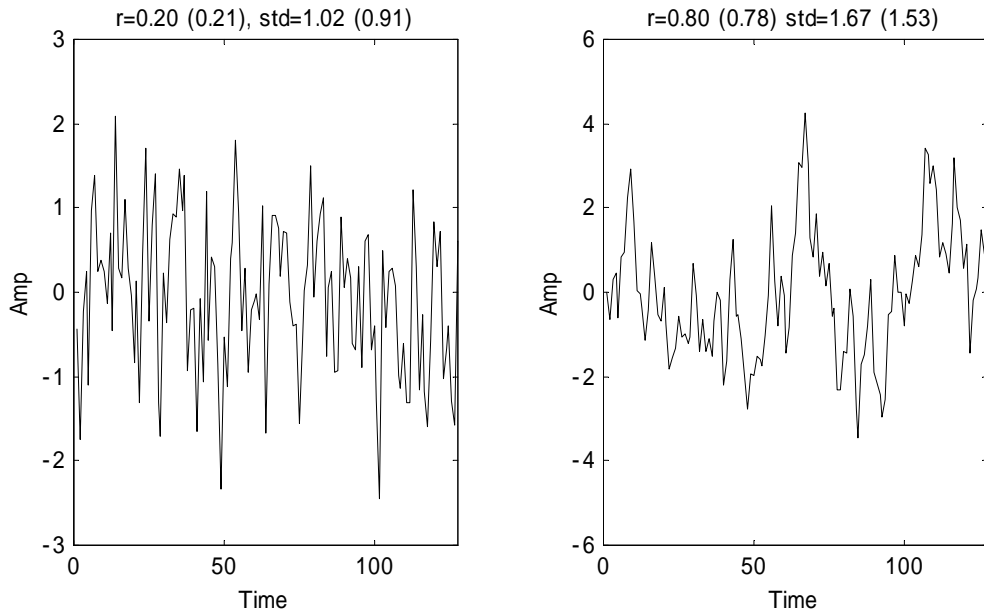


図 2-2. ラグ1自己相関係数 0.2 (左) と 0.8 (右) に対する，赤色ノイズモデルの時系列．測定されたラグ1自己相関係数はパネル上部のカッコ内に示している．やはりパネル上部に理論的な標準偏差と観測された標準偏差とを，観測された標準偏差にカッコを付けて示している．

赤色ノイズモデルを用いて，代理時系列を生成するには，通常(2.1)を用いる．この際，二つのパラメータ， $r$  と  $\sigma_w$  を知らなくてはならない．赤色ノイズの標準偏差を， $\sigma_R$  と表すことにすれば，

$$\begin{aligned} \sigma_R^2 &\equiv \langle y(t)^2 \rangle = r^2 \langle y(t-1)^2 \rangle + 2r\sigma_w \langle y(t-1)\varepsilon(t) \rangle + \sigma_w^2 \langle \varepsilon(t)^2 \rangle \\ &= r^2\sigma_R^2 + \sigma_w^2 \end{aligned}$$

であるから，

$$\sigma_w = \sqrt{(1-r^2)}\sigma_R \Leftrightarrow \sigma_R = \sigma_w / \sqrt{(1-r^2)} \tag{2.2}$$

となる．ここで $\langle \rangle$ は期待値（アンサンブル平均）．観測された時系列を赤色ノイズと仮定すれば，観測された分散が $\sigma_R^2$ であるので，上式から $\sigma_w$ が得られる．

多くの場合，ラグ1相関係数は，時系列から通常の相関係数の計算と同じように求められる．しかしこの方法で観測時系列からラグ1相関係数を求めると，求められたラグ1相関係数は若干小さく見積もられる，すなわちラグ1相関係数の推定が偏っていることが実は知られている．またラグ1相関係数の過小評価は(2.2)から白色ノイズ分散の過小評価につながり，これらのパラメータを使って行なう統計的なテストは有意となりやすくなる．この偏りはただしそれほど大きくなく，100個のデータから求める場合では0.02程度である．ラグ1相関係数の不偏推定方法は，Allen and Smith (1996)に示されているので，万一気難しい査読者に攻撃されたりして興味が湧いた読者は参照されたい．

なお，赤色ノイズモデルは，一次のマルコフ過程(first order Markov process)とも呼ばれる．マルコフ過程の定義は，未来の確率状態が過去によらずに決まる確立過程である．一次のマルコフ過

程は多分、1時間ステップ前の状態には依存するが、それよりも前の状態には依存しないということであろう。

### 2.2.2. 赤色ノイズの生成方法

赤色ノイズを生成するには、生成すべき時系列の長さを  $tlng$ 、時系列の個数を  $nsmpl$ 、ラグ1自己相関係数を  $r1$  とすると、例えば次のように生成することができる。

```
rtsrs=randn(tlng,nsmpl);           % tlng × nsmpl の白色ノイズを生成
rtsrs(1,:)=rtsrs(1,:)/sqrt(1-r1^2); % 最初の時間の赤色ノイズを規格化
for t=2:tlng
    rtsrs(t,:)=rtsrs(t-1,:)*r1+rtsrs(t,:);
end
```

各時系列の最初のデータは、白色ノイズを赤色ノイズと同じ標準偏差になるように規格化して与えている。

演習問題 2-1  $r1=0.5$ ,  $nsmpl=10000$ ,  $tlng=10$  で赤色ノイズを生成し、分散と標準偏差のヒストグラムを描け。また、分散の平均が理論値に一致し、標準偏差の平均は一致しないことを示せ。m 番目の

の時系列の分散を  $(1/n) \sum_{i=1}^n y_m(t)^2$  で求める。

回答

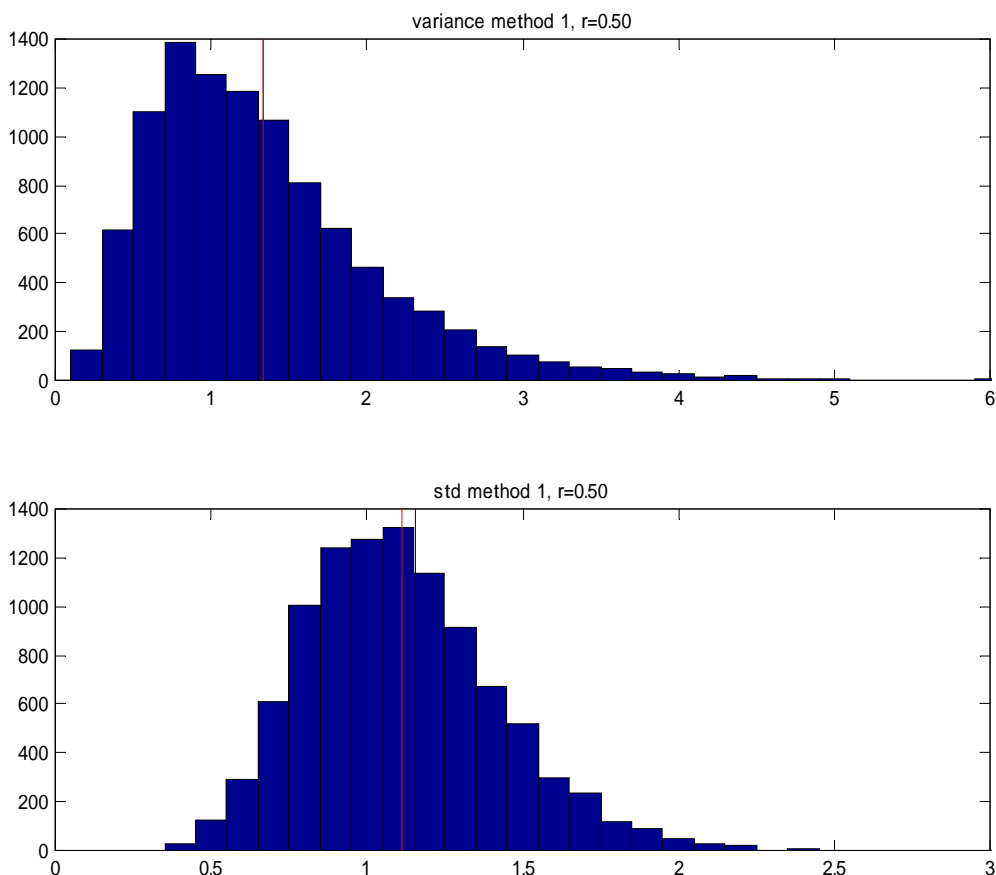


図 2-3 時系列の分散と標準偏差の分布図．各パネルの垂直線の黒は理論値，赤は時系列パラメータの平均値である．分散については両者がよく一致しているために，黒線が赤線に隠されている．

2.2.3. 自己回帰過程(autoregressive process)

赤色ノイズでは，ラグ 1 の情報の伝達のみを考慮したが，同様の考えで多くのラグを含めることができる．このようなより一般的なモデルを，**自己回帰モデル(autoregressive model)**と呼ぶ．

$p$  次の自己回帰過程は

$$y(t) + \phi(1)y(t-1) + \dots + \phi(p)y(t-p) + \sigma_w \varepsilon(t) = 0 \tag{2.3}$$

で表現される． $\phi(i)$  はモデルの係数である．時系列  $y$  がこの過程に従うことを， $y_t \sim \text{AR}(q)$  と書く．

**赤色ノイズモデルは 1 次の自己回帰モデル**である．係数  $\phi$  は AR1 ではラグ 1 相関係数に一致したが，2 次以上の AR モデルでは一致しない．なお，赤色ノイズモデルとは，各項の係数は逆の符号としている．これは Matlab の慣用にあわせており，両方の流儀がある．

自己回帰モデルがどのような振る舞いをするかは，平均的な振る舞いについてしか議論できないので，まずアンサンプル平均， $\bar{y}(t) = \langle y(t) \rangle$ ，を取って次式を得る．

$$\bar{y}(t) + \phi(1)\bar{y}(t-1) + \dots + \phi(p)\bar{y}(t-p) = 0$$

ここで，

$$\bar{y}(t) = \lambda^{t-p} \bar{y}(t-p) \tag{2.4}$$

を仮定すると (つまり  $\bar{y}(t) = \lambda \bar{y}(t-1)$  ),

$$\lambda^p + \phi(1)\lambda^{p-1} + \dots + \phi(p) = 0$$

という特性方程式と呼ばれる式を得る. 明らかに,  $|\lambda| > 1$  では  $\bar{y}$  が発散する解を持つし,  $\lambda = 1$  でもノイズが加わる効果によって平均が徐々にゼロからずれて分散が増加する. このような, 望ましくない特徴をもたないためには, 特性方程式の全ての根が  $|\lambda| < 1$  である必要がある. この場合, 遠く離れたデータの影響は減衰するので, 十分に初期値から離れれば初期値の影響は無視できる. この時 **AR 過程は定常**であると言う. AR1 では,  $|\phi(1)| < 1$  が定常の条件である. なお一般に, 確率過程のアンサンブル平均で得られる平均・分散・自己相関関数とが, 時間的に変化しないことを, 確率過程が**定常**<sup>1</sup>であるという.

#### 2.2.4. 自己回帰モデルの推定

自己回帰モデルの推定は, 統計的<sup>2</sup>予測の実行, この章で説明する自由度の推定, 次章で説明する Maximum Entropy Method (MEM) Spectrum を推定するなど, 様々な局面で用いられる. この節では, 自己回帰モデルの推定方法を説明しよう.

自己回帰モデルの係数  $\phi$  と, ラグ  $\tau$  の自己共分散関数 ( $\rho(\tau) \equiv \langle y(t)y(t+\tau) \rangle$ ) と関係は, 次の関係がある.

$$\begin{aligned} \rho(\tau) &\equiv \langle y(t)y(t-\tau) \rangle \\ &= -\langle \{ \phi(1)y(t-1) + \dots + \phi(p)y(t-p) + \sigma_w \varepsilon(t) \} y(t-\tau) \rangle \\ &= -\langle \phi(1)y(t-1)y(t-\tau) + \dots + \phi(p)y(t-p)y(t-\tau) + \sigma_w \varepsilon(t)y(t-\tau) \rangle \end{aligned} \tag{2.5}$$

ここで, ラグがゼロでなければ,

$$\begin{aligned} \rho(\tau) + \phi(1)\langle y(t-1)y(t-\tau) \rangle + \dots + \phi(p)\langle y(t-p)y(t-\tau) \rangle + \langle \sigma_w \varepsilon(t)y(t-\tau) \rangle &= 0 \\ \rho(\tau) + \phi(1)\rho(\tau-1) + \dots + \phi(p)\rho(\tau-p) &= 0 \end{aligned}$$

となる. 上の式を分散  $\sigma^2 = \rho(0)$  で割って, 自己共分散を自己相関  $r$  に変えれば,

$$r(\tau) + \phi(1)r(\tau-1) + \phi(2)r(\tau-2) + \dots + \phi(p)r(\tau-p) = 0 \tag{2.6}$$

が得られる. この関係から AR1 については,  $r(1) = -\phi(1)$  となるので, すでに述べたように AR1 のモデル係数はラグ 1 自己相関係数に他ならないことが確認できる. ただしラグ 0 については, (2.5)から落とした白色ノイズの項も含めることが必要であり, この場合

<sup>1</sup> 厳密には, この定常性は, 弱定常性あるいは共分散定常性であって, 強定常とは区別される. 強定常性は,  $y(t), t=1, \dots, n$  の分布が,  $y(t+\tau)$  の分布と一致することを言う. 正規分布に従う確率過程は, 弱定常であれば強定常でもある. この資料では定常とは常に弱定常の意味で使われる.

<sup>2</sup> 予測には統計的な方法と, 力学的なモデル計算による方法とがある.

$$\rho(0) + \phi(1)\langle y(t-1)y(t-\tau) \rangle + \dots + \phi(p)\langle y(t-p)y(t-\tau) \rangle + \langle \sigma_w \varepsilon(t)y(t) \rangle = 0$$

$$\therefore 1 + \phi(1)r(1) + \phi(2)r(2) + \dots + \phi(p)r(p) + \sigma_w^2 / \sigma^2 = 0 \tag{2.7}$$

である。自己相関は、対称関数（正負のラグで同じ値）であることを利用して、左辺のラグが1からpまでの(2.6)および(2.7)を行列式であらわせば、

$$\begin{bmatrix} r(1) & r(2) & \dots & r(p) \\ 1 & r(1) & \dots & r(p-1) \\ r(1) & r(2) & \dots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi(1) \\ \phi(2) \\ \vdots \\ \phi(p) \end{bmatrix} = \begin{bmatrix} 1 - \sigma_w^2 / \sigma^2 \\ r(1) \\ r(2) \\ \vdots \\ r(p) \end{bmatrix} \tag{2.8}$$

これはちょっと書き換えて、行列が Toeplitz(テプリッツ行列：成分が対角線に沿って一定である行列)になるようにできる。

$$\begin{bmatrix} 1 & r(1) & r(2) & \dots & r(p) \\ r(1) & 1 & r(1) & \dots & r(p-1) \\ r(2) & r(1) & 1 & \dots & r(p-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(p) & r(p-1) & r(p-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \phi(1) \\ \phi(2) \\ \vdots \\ \phi(p) \end{bmatrix} = \begin{bmatrix} -\sigma_w^2 / \sigma^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.9}$$

この関係式が有名な**ユール・ウォーカー方程式(Yule-Walker equation)**である。Matlab では `lpc`, `pyulear` で利用できる。

ユール・ウォーカー方程式で自己回帰モデルを得るには、次数 p を決めてやらなくてはならない。次数を大きくすればするほど、モデルの当てはまりがよくなるのは当然なので、それだけを考えて、真の p は小さいのにもかかわらず大きな p を選んでしまい勝ちである。この落とし穴を回避し、より妥当に p を決めるためのよく使われる方法としては、**赤池の情報量基準(Akaike's Information Criterion, AIC)**が最小になる次数 p を用いる方法である。AIC は自己回帰モデルに限らず広く使われる概念であって、その詳しい説明は柴田(1996)の7章を参照されたい。自己回帰モデルについての AIC は、

$$AIC = N + N \ln 2\pi\sigma_w^2 + 2(p+1) \tag{2.10}$$

であり、ここで N は時系列データの個数である。上式のうち自己回帰モデルに依存する部分のみ取り出して、

$$AIC\text{の变化部分} = N \ln \sigma_w^2 + (p+1) \tag{2.11}$$

として AIC を用いることもよく行なわれる。AIC を用いて AR モデルを推定する手続きは次の通りである。

- 1) あらかじめ定めた次数の範囲で、各次数についてユール・ウォーカー方程式から AR モデルの

係数を求める。

- 2) 各次数の AR モデルの AIC を求める。
- 3) 最小となる AIC を与える次数の AR モデルを採用する。

なお、ユール・ウォーカー方程式の最大のラグ相関係数  $\phi(p)$  も未知数であるとして、同時に決定するバーグ(Burg)のアルゴリズムは、ユール・ウォーカーのアルゴリズムより若干優れていると考えられている。バーグのアルゴリズムは、Matlab では **pburg** で利用できる。

後に述べるように、自己回帰モデルにもとづいてラグ相関係数を推定する応用もある。その場合、ラグが  $p$  以下であればラグ相関係数は観測された相関係数であり、ラグが  $p$  よりも大きいのであれば(2.6)を用いてラグ相関係数が求められる。

### 2.2.5. 移動平均(MA)過程

$q$  次の移動平均過程 (moving-average process, MA process)は

$$y(t) + \varepsilon(t) + \theta(1)\varepsilon(t-1) + \theta(2)\varepsilon(t-2) + \dots + \theta(q)\varepsilon(t-q), \quad t = 1, \dots, n$$

で表現され、時系列  $y$  がこの過程に従うことを、 $y_t \sim \text{MA}(q)$ と書く。

無限次 MA 過程は適当な条件のもとでは AR(1)で表現される。任意の有限次の MA 過程は常に定常である。

### 2.2.6. 自己回帰移動平均(ARMA)過程

$p, q$  次の自己回帰移動平均過程 (autoregressive-moving average process, ARMA process)を、 $y_t \sim \text{ARMA}(p, q)$ と書く(ハーベイ, 1985, 2章(1.25))

$$y(t) + \phi(1)y(t-1) + \dots + \phi(p)y(t-p) + \varepsilon(t) + \theta(1)\varepsilon(t-1) + \dots + \theta(q)\varepsilon(t-q) = 0, \quad t = 1, \dots, T$$

$p, q$  が共にゼロで無い場合に混合過程と呼ぶ。ARMA 過程の自己回帰部分が、AR 過程の定常性を満たす場合には、ARMA 過程もまた定常となる。

### 2.2.7. 自己回帰和分移動平均(ARIMA)過程

自己回帰和分移動平均 (autoregressive-integrated-moving average)過程、略して ARIMA 過程と呼ぶ。ARIMA 過程は、 $y(t) - y(t-1)$ が ARMA 過程となる過程である。ARIMA 過程は、局所的なトレンドを表現することができ、そのために一定の非定常性を表現することができる。

### 2.2.8. コレログラム

非負のラグに対する標本自己相関関数のプロットを**コレログラム(correlogram)**と呼ぶ。なお、自己相関関数とは、ラグの関数として複数の自己相関係数を扱う場合に用いる呼び名である。コレログラムはしばしば大気海洋の研究でも使われ、その場合主としてコレログラムに振動現象が見られるかどうか議論されることが多い。

## 2.3. 統計推定・検定の自由度

連続相関がある場合に、どのようにして推定・検定に用いる自由度を推定すればよいのかには、



自己相関関数を用いて推定する標準的な手法があるので、それを紹介しよう。基本的な考え方は、**実効的に独立な標本間の時間**(time between effectively independent samples, Trenberth (1984)および Metz (1991)の呼び名)、もしくは**有効無相関時間**(effective decorrelation time)と呼ばれる  $T_e$  で、データのサンプル数を割って、**有効自由度**(effective number of degrees of freedom)を得る。平均・分散・相関および回帰に関する有効無相関時間を下にまとめて示そう。

平均について 
$$T_e = \sum_{\tau=-N}^N \left(1 - \frac{|\tau|}{N}\right) r_{xx}(\tau) = 1 + 2 \sum_{\tau=1}^N \left(1 - \frac{\tau}{N}\right) r_{xx}(\tau) \quad (2.12)$$

分散について 
$$T_e = \sum_{\tau=-N}^N \left(1 - \frac{|\tau|}{N}\right) r_{xx}^2(\tau) = 1 + 2 \sum_{\tau=1}^N \left(1 - \frac{\tau}{N}\right) r_{xx}^2(\tau) \quad (2.13)$$

相関・回帰について 
$$T_e = \sum_{\tau=-N}^N \left(1 - \frac{|\tau|}{N}\right) [r_{xx}(\tau)r_{yy}(\tau) + r_{xy}(\tau)r_{yx}(\tau)] \quad (2.14)$$

$$= 1 + r_{xy}(0)r_{yx}(0) + 2 \sum_{\tau=1}^N \left(1 - \frac{\tau}{N}\right) [r_{xx}(\tau)r_{yy}(\tau) + r_{xy}(\tau)r_{yx}(\tau)]$$

ここで  $r$  は相関関数、 $N$  は標本の数（時系列の長さ）、 $\tau$  はラグである。相関関数は、共分散関数を全分散（つまり分母の和が  $1 \dots N$  である）で規格化して

$$r_{xx}(\tau) \equiv \frac{\frac{1}{N-\tau} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)}{\frac{1}{N} \sum_{t=1}^N x(t)^2}, r_{yy}(\tau) \equiv \frac{\frac{1}{N-\tau} \sum_{t=1}^{N-\tau} y(t)y(t+\tau)}{\frac{1}{N} \sum_{t=1}^N y(t)^2},$$

$$r_{xy}(\tau) \equiv \frac{\frac{1}{N-\tau} \sum_{t=1}^{N-\tau} x(t)y(t+\tau)}{\sqrt{\frac{1}{N} \sum_{t=1}^N x(t)^2 \frac{1}{N} \sum_{t=1}^N y(t)^2}}$$

定義される。注意すべきことは、**無相関時間は（従って有効自由度は）現象に内在する量ではなく、データの長さにも、またどの統計解析を行うかにも依存することである**。なお、上の式を使う場合の参考文献を上げておこう。どの式についても、Lieth (1982)で上げられているようだが、残念ながら元論文を入手できていない。平均については統計分野の Bayley and Hammersley (1946) や大気海洋では、Lieth (1982)や Trenberth (1984)が妥当なところである。相関は Bartlett が 1930 年代に定式化したらしくその内容は Bartlett(1955)にも記されている。相関について、大気海洋に導入したのは Davis (1976)である。分散は誰が出したのかよく分からない。Lieth は 1982 年の論文の前にもいくつか関連する話題について論文やレポートを書いているようだが(Trenberth 1984 の reference による)、どれも入手しづらい文献で彼の実際の貢献がどのようなものか判断できかねている。

なお、実際の計算に当たっては、大きなラグについては上式の分子の推定が使えるデータの数が少なく、良い推定が得られないという問題がある。そこで、Katz (1982)や Trenberth (1984)では、AR モデルを推定して、その AR モデルに基づいて自己相関係数を計算する方法を勧めている。こ

の方法は、ラグが大きなラグ相関の値が大きく、無相関時間の推定に影響が強い時に使うとよいだろう。もし、ラグが大きなラグ相関が小さいのであれば、AR モデルの推定を経由する必要はそれほどない。また Trenberth(1984)では、ラグが大きいと上式の分子の推定が使えるデータの数が少なくて悪くなるために、あえて偏りのある推定値、 $(1/N) \sum_{t=1}^{N-\tau} x(t)x(t+\tau)$  を使うことも提案している。

結局、有意性の検定を行うには、検定の種類に応じて、上のいずれかの無相関時間を求め、それから得られる有効自由度で各々の検定の自由度を置き換える。また無相関時間の推定には、標本相関係数をそのまま使うと危ない場合には AR モデルを経由して、相関係数を推定する。AR モデルの推定は、ユール・ウォーカー方程式で行い、AIC で次数を決める。というように、データが独立である統計の教科書では簡単な有意性検定も、独立ではない時系列解析では一気に大変になる。

2.3.1. 平均に関する無相関時間の導出

真の平均を  $\langle x \rangle$ ，標本平均を  $\{x\}$  とする。標本の2乗から真の平均の2乗を引くと，

$$v \equiv \langle (\{x\} - \langle x \rangle)^2 \rangle = \langle \{x\}^2 \rangle - 2\langle \{x\} \rangle \langle x \rangle + \langle x \rangle^2 = \langle \{x\}^2 \rangle - \langle x \rangle^2 \tag{2.15}$$

である。標本平均を定義に従って書き直して，

$$v = \left\langle \left( \frac{1}{N} \sum_{t=1}^N x(t) \right)^2 \right\rangle - \langle x \rangle^2 \tag{2.16}$$

であるので，ここで  $x$  を偏差と真の平均の和で表して， $x = x' + \mu$ ， $\mu = \langle x \rangle$ ，として

$$\begin{aligned} v &= \left\langle \left( \frac{1}{N} \sum_{t=1}^N x'(t) + \mu \right)^2 \right\rangle - \langle x \rangle^2 \\ &= \left\langle \frac{1}{N^2} \sum_{s=1}^N \sum_{t=1}^N (x'(t) + \mu)(x'(s) + \mu) \right\rangle - \mu^2 \\ &= \left\langle \frac{1}{N^2} \sum_{s=1}^N \sum_{t=1}^N (x'(t)x'(s) + (x'(s) + x'(t))\mu + \mu^2) \right\rangle - \mu^2 \\ &= \frac{1}{N^2} \sum_{s=1}^N \sum_{t=1}^N \left\{ \langle x'(t)x'(s) \rangle + \langle x'(s) + x'(t) \rangle \bar{x} + \langle \mu^2 \rangle \right\} - \mu^2 \\ &= \frac{1}{N^2} \sum_{s=1}^N \sum_{t=1}^N \langle x'(t)x'(s) \rangle \end{aligned} \tag{2.17}$$

である。ここで， $\tau = j - i$  が同じ値を持つものについての和という形にまとめれば，

$$v = \frac{1}{N^2} \sum_{\tau=-N}^N (N - |\tau|) \rho_{xx}(\tau) = \frac{\sigma^2}{N} \sum_{\tau=-N}^N \left( 1 - \frac{|\tau|}{N} \right) r_{xx}(\tau) \tag{2.18}$$

互いに無相関な正規分布する標準偏差が  $\sigma$  の分布を， $N_e$  点で平均すれば分散は

$$\sigma^2 / N_e = \sigma^2 T_e / N \tag{2.19}$$

である．ただし  $T_e$  は上で述べた有効無相関時間で  $T_e = N / N_e$  で定義される．(2.18)と(2.19)を比較すると，

$$T_e = \sum_{\tau=-N}^N \left(1 - \frac{|\tau|}{N}\right) r_{xx}(\tau) \tag{2.20}$$

となつて，これは，平均に関する有効無相関時間と自己相関関数との関係式(2.12)に他ならない．

なお，(2.17)の偏差の定義を標本平均からのずれとすると以下のようになつて，簡単な形は得られない．

$$\begin{aligned} v &= \left\langle \left( \frac{1}{N} \sum_{t=1}^N x'(t) + \bar{x} \right)^2 \right\rangle - \langle x \rangle^2 \\ &= \left\langle \frac{1}{N^2} \sum_{s=1}^N \sum_{t=1}^N (x'(t) + \bar{x})(x'(s) + \bar{x}) \right\rangle - \langle x \rangle^2 \\ &= \left\langle \frac{1}{N^2} \sum_{s=1}^N \sum_{t=1}^N (x'(t)x'(s) + (x'(s) + x'(t))\bar{x} + \bar{x}^2) \right\rangle - \langle x \rangle^2 \\ &= \frac{1}{N^2} \sum_{s=1}^N \sum_{t=1}^N \left\{ \langle x'(t)x'(s) \rangle + \langle x'(s) + x'(t) \rangle \bar{x} + \langle \bar{x}^2 \rangle \right\} - \langle x \rangle^2 \end{aligned} \tag{2.21}$$

### 2.3.2. 相関に関する無相関時間の導出

(2.14)式を導出してみよう． $\langle \rangle$ を期待値もしくは真の平均とし， $\{ \}$ を標本平均(sample mean)とする．すなわち，ある観測量  $q(i)$ ,  $i=1, \dots, N$  について  $\{q\} = (1/N) \sum_{n=1}^N q(i)$  である．二つの時系列  $x, y$  について，真の共分散を  $\langle xy \rangle = V$ ，sample variance を  $\{xy\} = V + v$  としよう．問題なのは，真の共分散と標本分散(sample variance) が一致せず，標本分散が通常大きいことである．

定義から， $v = \{xy\} - \langle xy \rangle$  であるから， $v^2 = \{xy\}^2 - 2\{xy\}\langle xy \rangle + \langle xy \rangle^2$  となる．ここで， $v^2$  の真の平均は，

$$\langle v^2 \rangle = \langle \{xy\}^2 \rangle - 2\langle \{xy\} \rangle \langle xy \rangle + \langle xy \rangle^2 = \langle \{xy\}^2 \rangle - \langle xy \rangle^2$$

で与えられる．最後の等号では， $\langle \{q\} \rangle = \langle q \rangle$  を用いた．さらに，この式を書きなおして，

$$\begin{aligned} \langle v^2 \rangle &= \left\langle \left( \frac{1}{N} \sum_{t=1}^N x(t)y(t) \right) \left( \frac{1}{N} \sum_{s=1}^N x(s)y(s) \right) \right\rangle - \langle xy \rangle^2 \\ &= \left\langle \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N x(t)y(t)x(s)y(s) \right\rangle - \langle xy \rangle^2 \\ &= \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N \langle x(t)y(t)x(s)y(s) \rangle - \langle xy \rangle^2 \end{aligned}$$

となる。ここで第一項の $\langle \rangle$ 内が結合正規分布するとすれば、結合正規分布する4つの確率変数の積の期待値は、次のように、二つの変数間の共分散の全ての組み合わせの和で表される(証明は例えば、川嶋・酒井, p.33)。すなわち、 $x_m, x_n, x_p, x_q$ を4つの結合正規分布する確率変数であるとすれば、

$$\langle x_m x_n x_p x_q \rangle \approx \langle x_m x_n \rangle \langle x_p x_q \rangle + \langle x_m x_p \rangle \langle x_n x_q \rangle + \langle x_m x_q \rangle \langle x_n x_p \rangle$$

であることを用いて、

$$\begin{aligned} \langle v^2 \rangle &= \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N \left[ \langle x(t)x(s) \rangle \langle y(t)y(s) \rangle + \langle x(t)y(s) \rangle \langle y(t)x(s) \rangle + \langle x(t)x(t) \rangle \langle y(s)y(s) \rangle \right] - \langle xy \rangle^2 \\ &= \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N \left[ \langle x(t)x(s) \rangle \langle y(t)y(s) \rangle + \langle x(t)y(s) \rangle \langle y(t)x(s) \rangle \right] \end{aligned}$$

となる。ここで、 $\tau = j - i$ が同じ値を持つものについての和という形にまとめれば、

$$\langle v^2 \rangle = \frac{1}{N} \sum_{\tau=-N}^N \left( 1 - \frac{|\tau|}{N} \right) \left[ \rho_{xx}(\tau) \rho_{yy}(\tau) + \rho_{xy}(\tau) \rho_{yx}(\tau) \right] \tag{2.22}$$

となる。ただし、 $\rho$ は次式で定義される自己共分散関数および相互共分散関数である。

$$\begin{aligned} \rho_{xx}(\tau) &= \langle x(t)x(t+\tau) \rangle, \rho_{yy}(\tau) = \langle y(t)y(t+\tau) \rangle, \\ \rho_{xy}(\tau) &= \langle x(t)y(t+\tau) \rangle, \rho_{yx}(\tau) = \langle y(t)x(t+\tau) \rangle \end{aligned}$$

である。さらに、意味を掴みやすい表現を得るために、 $x, y$ 各々の分散を $Q_x = \langle x^2 \rangle, Q_y = \langle y^2 \rangle$ と書いて、

$$r_{xx}(\tau) = \rho_{xx}(\tau) / Q_x, r_{yy}(\tau) = \rho_{yy}(\tau) / Q_y, r_{xy}(\tau) = \rho_{xy}(\tau) / \sqrt{Q_x Q_y}$$

と定義される相関係数を用いれば、(2.22)は、

$$\langle v^2 \rangle = \frac{Q_x Q_y}{N} \left\{ \sum_{\tau=-N}^N \left( 1 - \frac{|\tau|}{N} \right) \left[ r_{xx}(\tau) r_{yy}(\tau) + r_{xy}(\tau) r_{yx}(\tau) \right] \right\} \tag{2.23}$$

と表わされる。

一方、互いに無相関かつ自己相関を持たない正規分布する確率過程が与える、標本時系列

$x(t), y(t), t=1, \dots, N$  の共分散の期待値は,

$$\left\langle \frac{1}{N} \sum_{t=1}^N x(t)y(t) \right\rangle = \sqrt{\frac{Q_x Q_y}{N}}$$

である．そこで，上の共分散の増加分， $\langle v^2 \rangle$ ，に含まれる独立なデータの数を  $N_e$  (有効自由度，the effective number of degrees of freedom (effective sample size), Metz (1991))と定義すれば，

$$\langle v^2 \rangle = \frac{Q_x Q_y}{N} \left\{ \sum_{\tau=-N}^N \left( 1 - \frac{|\tau|}{N} \right) [r_{xx}(\tau)r_{yy}(\tau) + r_{xy}(\tau)r_{yx}(\tau)] \right\} = \frac{Q_x Q_y}{N_e} = \frac{Q_x Q_y}{N\Delta t / T_e}$$

と書くことができる．ここで， $T_e$  は有効無相関時間である．明らかに，

$$T_e = \Delta t \sum_{\tau=-N}^N \left( 1 - \frac{|\tau|}{N} \right) [r_{xx}(\tau)r_{yy}(\tau) + r_{xy}(\tau)r_{yx}(\tau)]$$

であり，これは(2.14)に他ならない．積分時間スケールから自由度は  $DOF = N\Delta t / T_e$  と求めることができる．上式の[]内第二項は通常小さいので，無視されることが多い．

なお，得られているデータ長  $N$  の範囲で，相関係数が十分に速くゼロになるならば，(2.14)の近似表現として，

$$T_e = \Delta t \sum_{\tau=-\infty}^{\infty} [r_{xx}(\tau)r_{yy}(\tau) + r_{xy}(\tau)r_{yx}(\tau)]$$

も用いられる．ここで， $\int_{-\infty}^{\infty}$  の和を取る範囲  $(-\infty, \infty)$  は実際的には，相関係数がゼロとみなせる範囲とすれば良い．この場合，積分時間スケールは，データを生成する物理過程によって定まる定数であると考えられることができる．

## 2.4. フィルタリング

時系列解析の様々な局面で，フィルタリングが用いられる．大きく分けると，フーリエ変換・逆変換を使う方法と，時間領域で重み付き移動平均を適用する方法 (非再帰的フィルター, Finite Impulse Response (FIR) フィルターとも呼ばれる)，時間領域で重み付き移動平均を再帰的に使う方法 (再帰的フィルター, Infinite Impulse Response (IIR) フィルターとも呼ばれる) の3つである．フーリエ変換・逆変換は通常のデータ解析であれば，まず考えるべき第一の方法であり，一般に FIR・IIR フィルターよりも良い特性をもつ (詳しくは Press ら(1993)を参照のこと)．時間領域の (非再帰的な) 重み付き移動平均の代表例は，重みを等しくした  $n$  点移動平均である．この移動平均はフィルターの性能はあまりよくないが，その単純さと広く使われているので大気海洋コミュニティで理解されやすいというなじみ深さの点から，これで済む程度であればこれを使うのは良い方法である．バターワース・フィルターなどの再帰的フィルターは，時間領域の非再帰的なフィルターに比べると，一般に高い性能を実現しやすい．一方で，位相がずれるなどの問題もある．また，フィルターの性能はフーリエ変換・逆変換に比べて高いわけではない．リアルタイムにデータを処理する必要があるならば FIR か IIR を使わねばならず，その場合には通常 IIR が用いられる．

Matlab ではフーリエ変換が `fft` , 逆変換が `ifft` である . 時間領域のフィルターは `filter` で , 特に位相のずれをなくしたい時は , 時間領域に順方向と逆方向の 2 回フィルターを通す `filtfilt` も用意されている . 様々なフィルターの設計もサポートされており , 例えば `butter` でバターワースフィルターが得られる .

## 参考文献

- Allen, Myles R., and Leonard A. Smith, 1996: Monte Carlo SSA: Detecting irregular oscillations in the Presence of Colored Noise. *J. Climate*, **9** (12), 3373–3404.
- Alexander , M. A., C. Deser, and M. S. Timlin, 1999: The Reemergence of SST anomalies in the North Pacific Ocean. *J. Climate*, **12** (8), 2419-2433.
- Bartlett, M. S., 1955: An introduction to stochastic processes with special reference to methods and applications. Cambridge university press, pp. 388. (third edition が 1978)
- Bayley G. V. and J. M. Hammersley, 1946: The 'effective' number of independent observations in autocorrelated time series. *J. Roy. Statist. Soc. Suppl.*, **8**(2), 184-197.
- Davis, R. E., 1976: Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.* **6** (3), 249-266.
- Hanan, E. J. 1970: Multiple Time Seies. Wiley.
- Katz, R. W., 1982: Statistical evaluation of climate experiments with genal circulation models: A parametric time series modeling approach. *J. Atmos. Sci.*, **39**, 446-1455. 1982.
- 川嶋 弘尚・酒井 英昭 : 現代スペクトル解析 . pp. 372, 森北出版株式会社 , 1989.
- Lieth, C. E., 1982: Statistical methods for the verification of long and short range forecasts. ECMWF seminar on Problems and prospects in long and medium range weather forecasting. 313-334. [available from ECMWF]
- Metz, W., 1991 : Optimal relationship of large-scale flow patterns and the barotropic feedback due to high-frequency eddies, *J. Atmos. Sci.*, **48**, 1141-1159.
- Namias, J., X., Yuan, and D. R. Cayan, 1988: Persistence of North Pacific Sea Surface Temperature and Atmospheric Flow Patterns. *J. Climate*, **1**, 682-703.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, 1993: Numerical Recipes in C [日本語版], 技術評論社, pp. 685.
- Trenberth K. E., 1984: Some effects of finite sample size and persistence on meteorological statistics. Part I: autocorrelations. *Mon. Wea. Rev.*, **112**, 2359-2368.