

周期時系列の統計解析

(1) 相関係数と位相差

nino

2017年11月3日

気温や水温の日間変動や季節変動などに代表される周期時系列を解析するために、多くの統計的手法が活用されている。はじめに、周期時系列の相関係数から位相差（時間のズレ）を求める方法について検討する。

コサイン関数の相関係数

まず、周期時系列をコサイン関数で近似し、その相関係数の式を求める。コサイン関数は $X_i = \cos \theta_i$ と $Y_i = \cos(\theta_i + \varphi)$ の2変数である。ここで、周期を T 、位相を φ 、離散変数を $\theta_i = 2\pi(i/T)$ 、そしてデータ数は整数周期分すなわち $n = lT$ とすると、次式が成り立つ。（ここで、 $l = 1, 2, 3, \dots$ の正整数）

$$\sum_{i=1}^n \cos \theta_i = \sum_{i=1}^n \cos(\theta_i + \varphi) = 0$$

変数 X_i , Y_i の総和は0となるから、それらの平均値は0である。

$$\bar{X} = \bar{Y} = 0$$

相関係数 R_{XY} は X_i と Y_i の共分散 S_{XY} および標準偏差 S_X , S_Y を用いると次式で表される。

$$R_{XY} = \frac{S_{XY}}{S_X S_Y}$$

変数 X_i , Y_i の平均値はともに0であることを用いて、右辺の各項を整理すると、

$$S_{XY} = \frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \frac{1}{n} \sum X_i Y_i$$

$$S_X = \sqrt{\frac{1}{n} \sum (X_i - \bar{X})^2} = \sqrt{\frac{1}{n} \sum X_i^2} \quad S_Y = \sqrt{\frac{1}{n} \sum (Y_i - \bar{Y})^2} = \sqrt{\frac{1}{n} \sum Y_i^2}$$

が得られる。これら共分散と標準偏差について、積和公式と倍角公式を用いてまとめる。ただし、倍角の総和： $\sum \cos 2\theta_i$ と $\sum \cos 2(\theta_i + \varphi)$ も0となることを利用した。

$$\begin{aligned} S_{XY} &= \frac{1}{n} \sum \cos \theta_i \cos(\theta_i + \varphi) \\ &= \frac{1}{n} \sum \frac{1}{2} \{ \cos(2\theta_i + \varphi) + \cos(-\varphi) \} \\ &= \frac{1}{2n} \sum \cos \varphi \\ &= \frac{1}{2} \cos \varphi \end{aligned}$$

$$\begin{aligned}
S_X &= \sqrt{\frac{1}{n} \sum \cos^2 \theta_i} \\
&= \sqrt{\frac{1}{n} \sum \left(\frac{1 + \cos 2\theta_i}{2} \right)} \\
&= \sqrt{\frac{1}{n} \sum \left(\frac{1}{2} \right)} \\
&= \sqrt{\frac{1}{2}}
\end{aligned}$$

同様に,

$$S_Y = \sqrt{\frac{1}{n} \sum \cos^2(\theta_i + \varphi)} = \sqrt{\frac{1}{2}}$$

したがって、相関係数 R_{XY} は

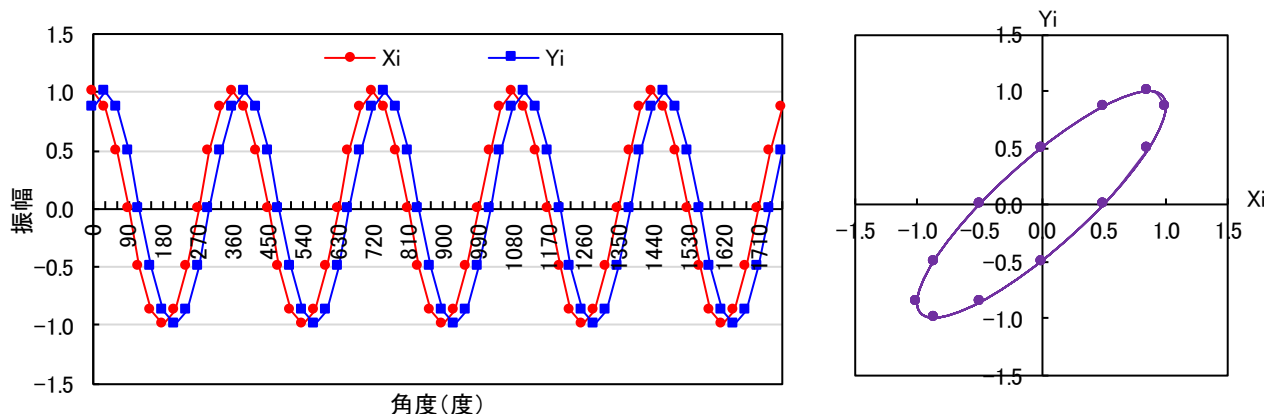
$$R_{XY} = \frac{S_{XY}}{S_X S_Y} = \cos \varphi$$

で表される。このように、相関係数は2変数の位相差のコサインで表される。言い換えると、相関係数のアークコサインから位相差を求めることができる。例えば、相関係数が0.866の時、位相差は $\arccos(0.866) = \pi/6$ すなわち30度が得られる。

なお、コサイン関数をサイン関数としても $R_{XY} = \cos \varphi$ となり、同じ結果が得られる。

例として、時系列図は周期： $T=360$ 度、位相： $\varphi=-30$ 度、期間：5周期分、そしてサンプリング間隔：30度のデータである。 Y_i は X_i より30度遅れており、相関係数は $\cos \varphi = 0.866$ となる。気温の年変化と想定すると、30度は12か月 \times 30度/360度=1か月に相当する。

散布図では、両変数の関係は楕円（長軸： $a=1.366$ 、短軸： $b=0.366$ 、 $b/a=0.268$ ）を左45度回転させた分布を示し、同一楕円上を左回りで移動している。相関係数は長軸と短軸の比 b/a を用いると、 $R = \{1 - (b/a)^2\} / \{1 + (b/a)^2\}$ で表される。詳細は、本サイト「寄稿」の「楕円の相関係数（その1）」を参照。この式に、 $b/a=0.268$ を代入すると、 $R_{XY}=0.866$ となり、相関係数と一致する。このように時系列と散布図は対応している。



トレンドを含む周期時系列のコサイン関数モデル

実際の時系列，例えば，気温の年変化では周期変動以外に，傾向変動（トレンド）や循環変動などの成分が含まれる．ここでは，トレンドを含む周期時系列について検討する．

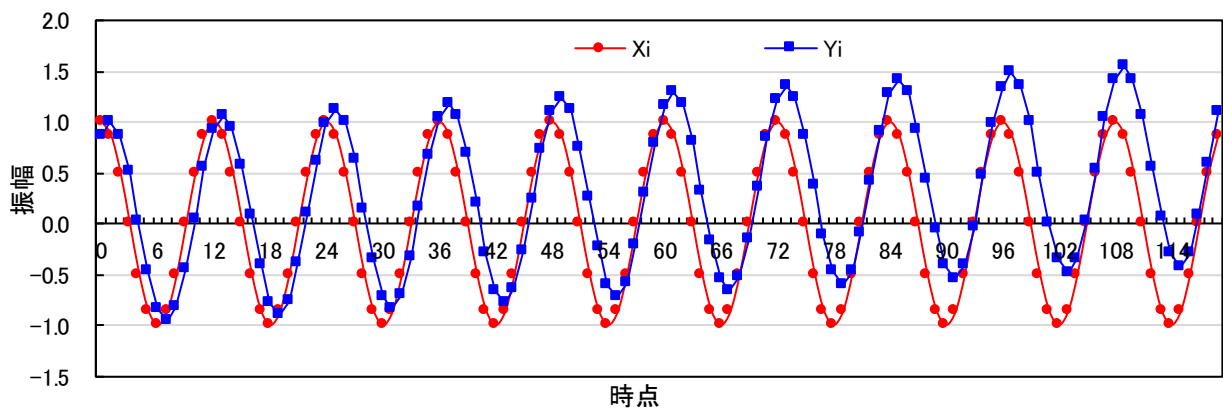
周期時系列は次の2つのコサイン関数モデル： $X_i = H_x \cdot \cos \theta_i + i \cdot k_x$ と $Y_i = H_y \cdot \cos(\theta_i + \varphi) + i \cdot k_y$ とした．ここで， H_x と H_y は振幅、 k_x と k_y はトレンドで，単位時点*i* 当たりの変化量である．両式をそれぞれ H_x ， H_y で除して， $X_i/H_x = \cos \theta_i + i \cdot k_x/H_x$ ， $Y_i/H_y = \cos(\theta_i + \varphi) + i \cdot k_y/H_y$ とし，あらためてトレンドを $k_x/H_x = K_x$ ， $k_y/H_y = K_y$ ，すなわち振幅 H_x ， H_y に対するトレンド k_x ， k_y の比率として定義し直した．以下， K_x ， K_y を相対トレンドという．したがって，コサイン関数モデルは， $X_i/H_x = \cos \theta_i + i \cdot K_x$ と $Y_i/H_y = \cos(\theta_i + \varphi) + i \cdot K_y$ で表される．この2式は，いずれも振幅が1で平均が0の周期変動に，相対トレンドの項が加わった式である．このように式を変換しても2変数の相関係数は変わらず，振幅の違いによる相関係数への影響が無視できるので，解釈がしやすくなる．

この2変数 X_i ， Y_i の相関係数 $R_{XY} = S_{XY}/S_X S_Y$ を計算する過程はやや複雑なので，参考までに結果だけを記した．

$$R_{XY} = \frac{\frac{1}{2} \left\{ \cos \varphi + K_Y + K_X \sin(\pi/T + \varphi) / \sin(\pi/T) + (n^2 - 1) K_X K_Y / 6 \right\}}{\sqrt{\frac{1}{2} \left\{ 1 + 2K_X + (n^2 - 1) K_X^2 / 6 \right\}} \sqrt{\frac{1}{2} \left\{ 1 + 2K_Y \sin(\pi/T + \varphi) / \sin(\pi/T) + (n^2 - 1) K_Y^2 / 6 \right\}}}$$

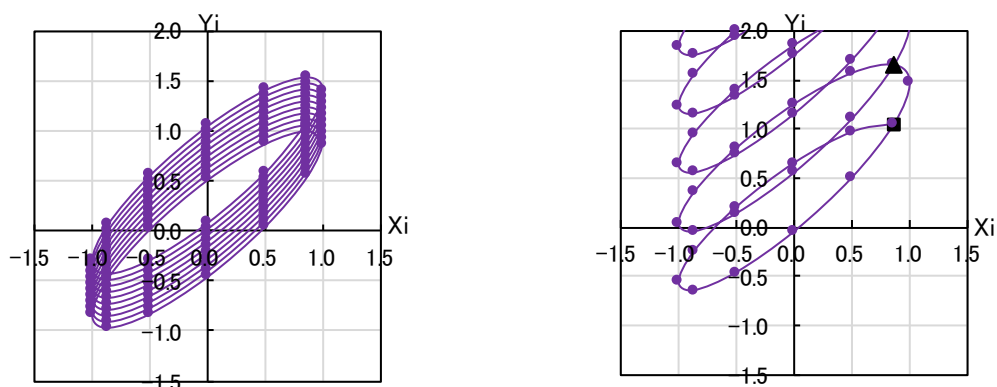
この理論式によると，相対トレンドを含む相関係数は，位相差 φ ，周期 T ，データ数 n ，および相対トレンド K_x ， K_y の5つのパラメーターを含んでいる．これらパラメーターによる相関係数への影響を検討するのは煩雑である．そこで，次節で具体例として取り上げる太陽高度と気温の年変化における位相差に対応したコサイン関数モデルを作成し検討する．なお，両変数とも相対トレンドが無い場合，すなわち $K_x = 0$ ， $K_y = 0$ を理論式に代入すると， $R_{XY} = \cos \varphi$ が得られ，先の結果と一致する．

時系列図は，太陽高度と気温の年変化を想定して，周期 $T = 12$ （か月），期間 = 10周期（年），変数 Y_i の位相差 $\varphi = -\pi/6$ （-1か月），相対トレンド $K_y = 0.005$ （/か月）の場合である．

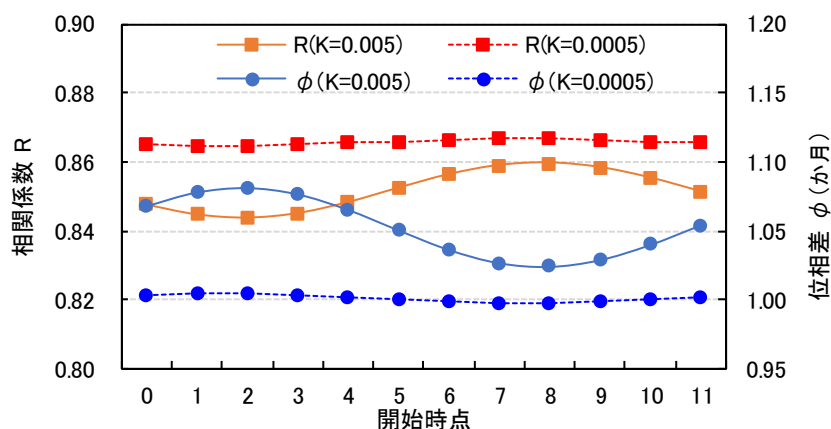


変数 Y_i は経時的に上昇していることが認められる. 相関係数は相対トレンドが無い場合の0.866よりも小さい0.835となった. そのため, 位相差は約1.11か月となり, 約0.11(11%)増加した.

一方, $K_Y=0.005$ の場合の $X_i Y_i$ 散布図(左図)をみると, 楕円が層状に重なっているように見られる. その状態をわかりやすくするため, 右散布図には, 相対トレンドを10倍にした $K_Y=0.05$ の場合を示した. 開始時点の四角記号■から三角記号▲までが1周期に相当し, 相対トレンドが0の場合は両記号は一致し閉じた楕円を形成するが, 相対トレンドがあるため, その相対トレンド分ずつY軸方向にらせん状に移動していく. このように, 周期時系列に相対トレンドが含まれると, 散布図では, らせん状の楕円が形成される.



このことから, 開始時点(月)が変わることによってらせんの形状が少し変わり, 相関係数と位相差に影響することが示唆される. この開始時点の違いによる相関係数と位相差への影響の程度を把握するため, 変数 Y_i の相対トレンドが $K_Y=0.0005$ と $K_Y=0.005$ の場合について, 開始時点が $i=0\sim 11$ までの相関係数 R_{XY} と位相差 ϕ の変化を調べた.(下図)



$K_Y=0.0005$ の場合は, 相関係数は0.865~0.867, 位相差は0.997~1.005か月の範囲にあった. 相対トレンドが無い場合の理論値は相関係数が0.866, 位相差は1.00か月であり, 開始時点の違いによる相関係数への影響はほとんどない. しかし, $K_Y=0.005$ の場合は, 相関係数は0.832~0.850, 位相差は1.06~1.12か月の範囲にあり, 位相差は理論値より6~12%増加した. また, 位相差が小数点2ケタを四捨五入して1.0か月となる相対トレ

ドを許容すると、位相差が1.00～1.03か月の範囲を示す $K_y=0.002$ が得られた。したがって、相対トレンドが0.002程度までならば、位相差はほぼ理論値に近いと判断できる。なお、位相差は対象期間や相対トレンド K_x 等によっても変化するので、この方法を実データに適用する場合は、それに類似のモデルを構築して検討し確認しておく必要がある。

一方、先述の相関係数の理論式には開始時点に関連した定数項が含まれていないが、開始時点 $i=6$ の時における相関係数と位相差が理論式の値と一致する。この開始時点 $i=6$ は、コサイン関数の値が-1（最低値）をとる $\theta_i=\pi$ の時点に相当する。そのため、理論式を用いる場合は開始時点に注意する必要がある。

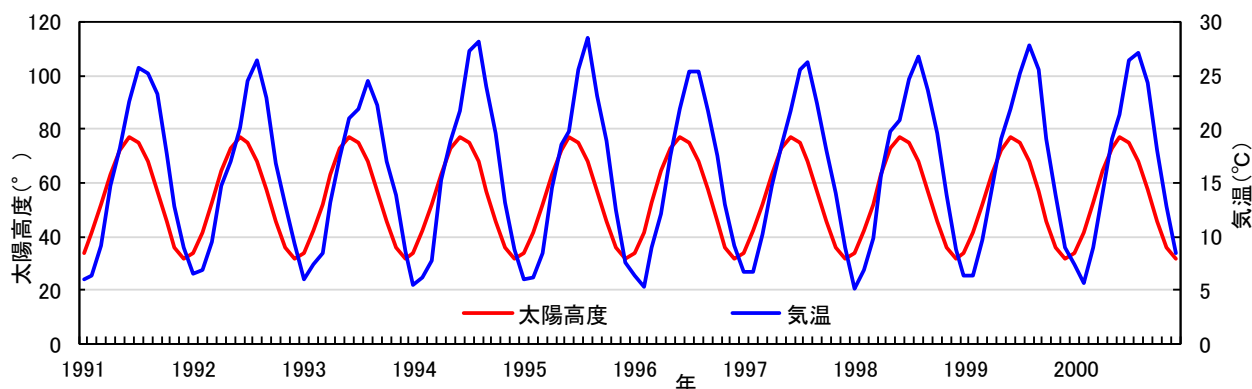
太陽高度と気温の位相差

相関係数から位相差を求める具体例として、気温と太陽高度について検討する。

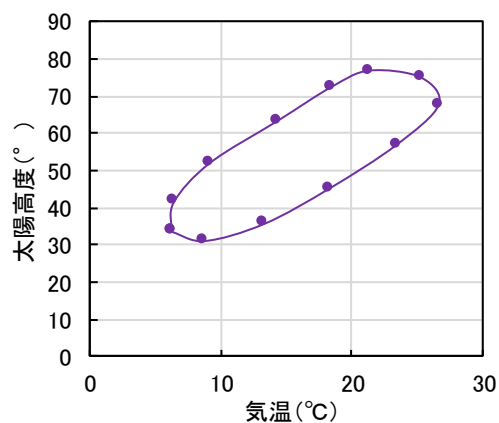
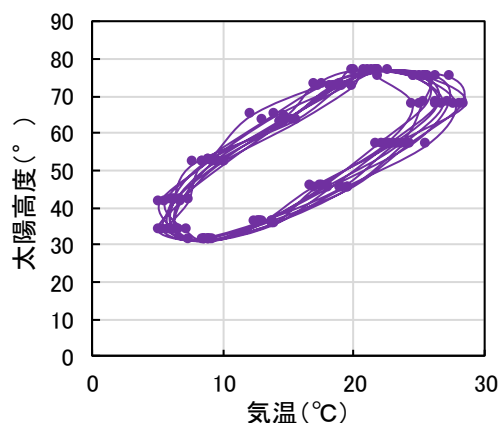
太陽光はまず地面を暖め、その熱が空気に伝わって空気を暖めるため、太陽高度と気温のピークに時間のズレが生じる。具体的には、1年間で太陽高度が最も高くなるのは夏至だが、最も暑いのは8月頃であり、気温の最高値は太陽高度よりも時間遅れで現れる。

用いたデータは、Casioのkeisanサービスの太陽高度（同一時刻の年間変化）と気象庁の月平均気温（横浜）である。太陽高度については、時刻が12時の5日間隔のデータから月平均値に変換した。対象期間は、1991年1月から2000年12月までの10年間である。また、振幅と平均は、太陽高度で 23° と 55° 、気温で 10°C と 16°C であった。

まず、時系列図をみると、太陽高度と気温はともにコサイン関数に類似した変動を示すが、気温は気象の影響を受け年毎に高低が認められた。また、気温のピークは太陽高度のピークよりも時間遅れが生じている。



次に、左散布図によると、少し歪んだ楕円状の分布を示した。別途、トレンドを推定したところ、太陽高度のトレンドは 0.000° /月、気温のそれは約 0.006°C /月（相対トレンド K_y として 0.0006 /月）であった。したがって、両者のトレンドは相関係数や位相差にほとんど影響を与えないと考えられた。この10年間における太陽高度と気温の相関係数は0.763、位相差は-1.34月となり、気温のピークは太陽高度よりも約1.3か月遅れと推察された。



一方、右散布図は10年間平均値を用いた場合であり、ほぼ楕円に近い形状を示した。この場合の位相差は -1.32 か月となり、約1.3か月で先の結果と一致した。しかし、年別の位相差を求めたところ、位相差は $-1.22 \sim -1.38$ か月の範囲にあり、年毎に少し違いが認められた。これは、主として年毎の気象変化にともなう気温変動の違いが影響したためである。相関係数から位相差を求めるためには、1年間のデータを用いるのは極力避け、複数年あるいはその平均を用いた方が良くと考えられた。

このように、太陽高度と気温の相関係数から位相差を求めることができるのは、両者が比較的コサイン関数に類似し、トレンド等が相関係数にほとんど影響を及ぼさないからである。相関係数から位相差を求める対象は限られると考えられるが、トレンド等の影響が無視できることを確認できれば、簡便で利用しやすい方法である。

参考文献)

- 1) Casioのkeisanサービス <http://keisan.casio.jp/exec/system/1185782617>
- 2) 気象庁 <http://www.data.jma.go.jp/obd/stats/etrn/index.php>
- 3) 物理のかぎしっぽ>楕円の相関係数(その1)
http://hooktail.sub.jp/contributions/ellipse_cor_01v4.pdf