楕円の相関係数（その１）

nino

2017 年 6 月 10 日

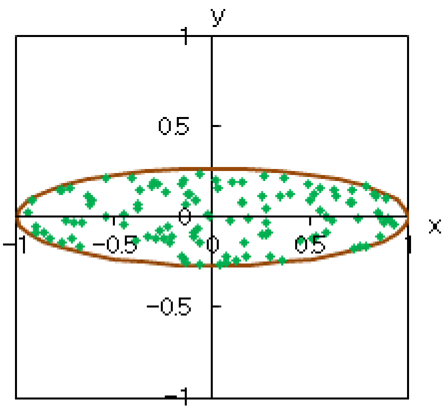
相関係数は様々な分野で使われていますが，その数値の解釈は漠然としたものです．それを少しでもわかりやすくできないかと考え，楕円の相関係数を求めてみました．結論を先に示すと，楕円（長軸：a、短軸：b）の相関係数 R は $R=(a^2-b^2)/(a^2+b^2)$ となりました．

楕円の直交座標

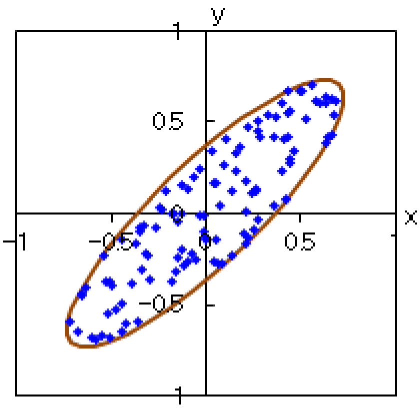
楕円方程式は次式で表されます．

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

例として，下図は，楕円（長軸：a=1，短軸：b =0.268）とその楕円内に生成させた乱数（112個のデータ）の分布です．乱数分布の相関係数は0.013となります．



次の図は，前図の楕円と乱数分布を左45度回転させたもので，乱数分布の相関係数は0.873となります．一般的に，このような分布について，相関係数を求める場合が多いようです．



楕円の左45度回転および長軸と短軸の長さが相関係数の値に影響していることが示唆されるので，回転楕円の方程式から相関係数を導出してみます．

回転楕円の媒介変数

楕円の媒介変数は，$x=a\cos\theta$，$y=b\sin\theta$ で表されます．

したがって，楕円(x,y)を左回りにα度回転させた楕円(X,Y)は，

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

ですから，α＝45°の場合は α＝π/4 を代入すると，

$$X = 1/\sqrt{2}(a\cos\theta - b\sin\theta)$$
$$Y = 1/\sqrt{2}(a\cos\theta + b\sin\theta)$$

となります．

一方，三角関数においては，θが 1 周期分すなわち 0 から 2π の範囲を考えればよいので，離散変数を $\theta_i = 2\pi\,(i/\mathrm{n})$とすると，

$$\sum_{i=1}^{n}\cos\theta_i = \sum_{i=1}^{n}\sin\theta_i = 0$$

が成り立ちます．Σは以下同様とします。したがって，変数 X_i，Y_i の総和およびそれらの平均値はいずれも 0 となります．

$$\sum X_i = \sum Y_i = 0 \qquad \overline{X} = \overline{Y} = 0$$

回転楕円の相関係数

45 度回転楕円の相関係数 R_{XY} を X_i と Y_i の共分散 S_{XY} および標準偏差 S_X，S_Y で表すと，

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

となります．まず，変数 X_i，Y_i の平均値はいずれも 0 であることを用いて，右辺の各変数を整理すると，

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

$$S_X = \sqrt{\frac{1}{n}\sum\left(X_i - \overline{X}\right)^2} = \sqrt{\frac{1}{n}\sum Xi^2}$$

$$S_Y = \sqrt{\frac{1}{n}\sum\left(Y_i - \overline{Y}\right)^2} = \sqrt{\frac{1}{n}\sum Y_i^2}$$

が得られます．次に，それぞれの式に X_i, Y_i を代入し，倍角公式と積和公式を用いてまとめると，

$$
\begin{aligned}
S_{XY} &= \frac{1}{n}\sum X_i Y_i \\
&= \frac{1}{2n}\sum (a\cos\theta_i - b\sin\theta_i)(a\cos\theta_i + b\sin\theta_i) \\
&= \frac{1}{2n}\sum \left(a^2\cos^2\theta_i - b^2\sin^2\theta_i\right) \\
&= \frac{1}{2n}\sum \left\{a^2\left(\frac{1+\cos 2\theta_i}{2}\right) - b^2\left(\frac{1-\cos 2\theta_i}{2}\right)\right\} \\
&= \frac{1}{4n}\sum \left(a^2 - b^2\right) \\
&= \frac{a^2-b^2}{4}
\end{aligned}
$$

$$
\begin{aligned}
S_X &= \sqrt{\frac{1}{n}\sum X_i^{\,2}} \\
&= \sqrt{\frac{1}{n}\sum \left(\frac{a\cos\theta_i - b\sin\theta_i}{\sqrt{2}}\right)^2} \\
&= \sqrt{\frac{1}{2n}\sum \left(a^2\cos^2\theta_i - 2ab\cos\theta_i\sin\theta_i + b^2\sin^2\theta_i\right)} \\
&= \sqrt{\frac{1}{2n}\sum \left\{\frac{\left(a^2+b^2\right)}{2} - 2ab\left(\frac{\sin 2\theta_i - \sin 0}{2}\right)\right\}} \\
&= \sqrt{\frac{1}{2n}\sum \frac{\left(a^2+b^2\right)}{2}} \\
&= \frac{\sqrt{a^2+b^2}}{2}
\end{aligned}
$$

同様に

$$
\begin{aligned}
S_Y &= \sqrt{\frac{1}{n}\sum Y_i^{\,2}} \\
&= \frac{\sqrt{a^2+b^2}}{2}
\end{aligned}
$$

となります．したがって，回転楕円の相関係数 R_{XY} は，

$$
\begin{aligned}
R_{XY} &= \frac{S_{XY}}{S_X S_Y} \\
&= \frac{a^2-b^2}{a^2+b^2} \\
&= \frac{1-(b/a)^2}{1+(b/a)^2}
\end{aligned}
$$

となります.

このように回転楕円の相関係数は長軸と短軸の長さのみで表され，それらの比 b/a が同じならば相関係数は等しくなります．例えば，回転楕円内の乱数分布の相関係数は 0.873 でしたが，回転楕円の相関係数は式に a=1 と b=0.268 を代入すると 0.866 となり，乱数分布の相関係数に近い値が得られました.
回転なしの楕円の場合については，S_{XY}＝0 ですから，その相関係数は 0 となります．最初に示した乱数分布の相関係数は 0.013 であり，ほぼ 0 に近い値となっています．このように，回転なしの楕円も回転楕円と同じ長軸と短軸の長さをもっているのに，違う結果が得られました．このことについては，今後検証していく予定です.
一方，相関係数 R_{XY} を変数として長軸と短軸の長さの比 b/a を表すと，$a \geqq b \geqq 0$ の条件では，

$$\frac{b}{a} = \sqrt{\frac{1 - R_{XY}}{1 + R_{XY}}}$$

となります．例えば，この式に回転楕円内の乱数分布の相関係数 0.873 を代入すると，b/a =0.260 が得られ，楕円の b 値（0.268）とほぼ一致しました．b/a 比は y 軸方向の変動幅を示していますので，相関関係を評価する上でよりわかりやすい指標となっています.