Correlation Coefficient of Ellipse (Section 1)

Nino

June 4, 2019

My name is Nino (pen name). This article is an English version of the article I wrote (URL below: Japanese version).
http://hooktail.sub.jp/contributions/ellipse_cor_01v4.pdf

Introduction

How should we evaluate the correlation coefficient? The correlation seems to be related to an ellipse, so I calculated the correlation coefficient of the ellipse. In conclusion, it became clear that the correlation coefficient r of the ellipse (major radius: a, minor radius: b) is given by $r=(a^2-b^2)/(a^2+b^2)$.

Cartesian coordinates of an ellipse

The equation of an ellipse is given by

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

Where : a is the major radius , b is the minor radius.

As an example, an ellipse ($a=1$, $b=0.268$) and a dataset of random numbers generated in the ellipse are shown in Figure 1.

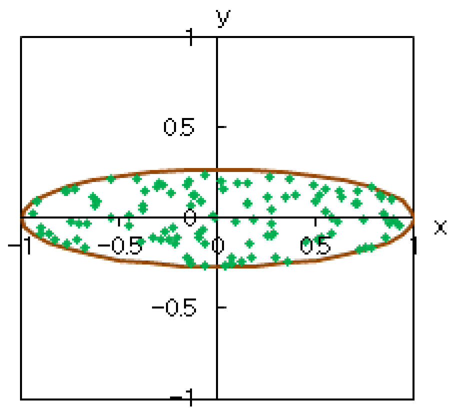

Fig.1 Scatter plot of the ellipse ($a=1$, $b=0.268$) and the random dataset(112 data) generated in the ellipse

The correlation coefficients of the ellipse and the random dataset are zero and 0.013 respectively. Figure 2 shows both the ellipse and the random dataset in Figure 1 rotated 45 degrees counterclockwise around the origin.

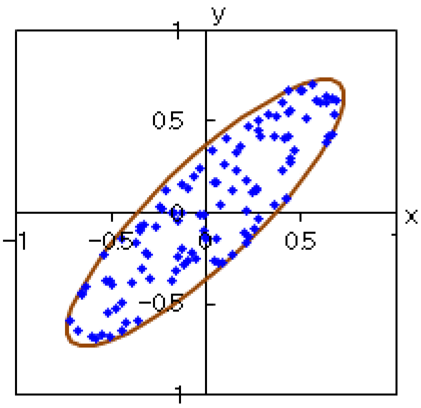

Fig.2 Scatter plot of the ellipse and the random dataset in Fig.1 rotated 45° counterclockwise

The correlation coefficients of the rotated ellipse and the random dataset are 0.866 and 0.873 respectively. In general, we often calculate the correlation coefficient of such a random dataset distribution. It is suggested that the rotation of the ellipse affects the correlation coefficient, so I tried to derive the correlation coefficient of the rotated ellipse.

The parametric formula of an ellipse is given by

$$x = a\cos\theta \qquad (2)$$
$$y = b\sin\theta$$

Where:

a the major radius ,and b is the minor radius.

θ is the parameter, which ranges from 0° to 360° .

A new coordinate (X, Y) obtained by rotating the coordinate (x, y) by α degrees counterclockwise around the origin is given by the following equation.

$$X = x\cos\alpha - y\sin\alpha \qquad (3)$$
$$Y = x\sin\alpha + y\cos\alpha$$

Substituting α = 45° into Eq.(3) gives the following equation.

$$X = 1/\sqrt{2}(a\cos\theta - b\sin\theta) \qquad (4)$$
$$Y = 1/\sqrt{2}(a\cos\theta + b\sin\theta)$$

Time series data are usually discrete values. In Eq.(2), assuming that the total number of samples in one cycle is n, the angle of the ith sample is represented by θ_i = 360 (i / n). The sum of $\cos\theta_i$ and the sum of $\sin\theta_i$ in one cycle are both zero. Therefore, the mean $\overline{X}$ of the Xi variable and the mean $\overline{Y}$ of the Yi variable are both zero.

$$\sum_{i=1}^{n}\cos\theta_i = \sum_{i=1}^{n}\sin\theta_i = 0 \tag{5}$$

$$\overline{X} = \overline{Y} = 0 \tag{6}$$

Correlation coefficient of an ellipse rotated 45° counterclockwise

The correlation coefficient R_{XY} is given by the following equation using the covariance S_{XY} and the standard deviations S_X and S_Y.

$$R_{XY} = \frac{S_{XY}}{S_X S_Y} \tag{7}$$

First, using the fact that the average values of the variables X_i and Y_i are both zero, the terms on the right side of the Eq.(6) are rearranged.

$$S_{XY} = \frac{1}{n}\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = \frac{1}{n}\sum X_i Y_i \tag{8}$$

$$S_X = \sqrt{\frac{1}{n}\sum\left(X_i - \overline{X}\right)^2} = \sqrt{\frac{1}{n}\sum Xi^2} \tag{9}$$

$$S_Y = \sqrt{\frac{1}{n}\sum\left(Y_i - \overline{Y}\right)^2} = \sqrt{\frac{1}{n}\sum Y_i^2} \tag{10}$$

Next, after substituting X_i and Y_i in Eq.(3) into Eq.(7), (8) and (9), the following equations are obtained by using trigonometric identities(1).

$$\begin{aligned} S_{XY} &= \frac{1}{n}\sum X_i Y_i \\ &= \frac{1}{2n}\sum\left(a\cos\theta_i - b\sin\theta_i\right)\left(a\cos\theta_i + b\sin\theta_i\right) \\ &= \frac{1}{2n}\sum\left(a^2\cos^2\theta_i - b^2\sin^2\theta_i\right) \\ &= \frac{1}{2n}\sum\left\{a^2\left(\frac{1+\cos 2\theta_i}{2}\right) - b^2\left(\frac{1-\cos 2\theta_i}{2}\right)\right\} \\ &= \frac{1}{4n}\sum\left(a^2 - b^2\right) \\ &= \frac{a^2 - b^2}{4} \end{aligned} \tag{11}$$

$$
\begin{aligned}
S_X &= \sqrt{\frac{1}{n}\sum X_i^{\,2}} \\
&= \sqrt{\frac{1}{n}\sum\left(\frac{a\cos\theta_i - b\sin\theta_i}{\sqrt{2}}\right)^2} \\
&= \sqrt{\frac{1}{2n}\sum\left(a^2\cos^2\theta_i - 2ab\cos\theta_i\sin\theta_i + b^2\sin^2\theta_i\right)} \\
&= \sqrt{\frac{1}{2n}\sum\left\{\frac{\left(a^2+b^2\right)}{2} - 2ab\left(\frac{\sin 2\theta_i - \sin 0}{2}\right)\right\}} \\
&= \sqrt{\frac{1}{2n}\sum\frac{\left(a^2+b^2\right)}{2}} \\
&= \frac{\sqrt{a^2+b^2}}{2}
\end{aligned}
\qquad (12)
$$

$$
\begin{aligned}
S_Y &= \sqrt{\frac{1}{n}\sum Y_i^{\,2}} \\
&= \frac{\sqrt{a^2+b^2}}{2}
\end{aligned}
\qquad (13)
$$

From these three equations, the correlation coefficient R_{XY} is given by

$$
\begin{aligned}
R_{XY} &= \frac{S_{XY}}{S_X S_Y} \\
&= \frac{a^2-b^2}{a^2+b^2} \\
&= \frac{1-(b/a)^2}{1+(b/a)^2}
\end{aligned}
\qquad (14)
$$

As a conclusion, it became clear that the correlation coefficient of the rotated ellipse is represented by the major and minor radius, and that the correlation coefficients between variables with the same *b/a* value are equal.

Substituting $a=1$ and $b=0.268$ into the Eq.(14), the correlation coefficient of the ellipse R_{XY} is 0.866, which matches the correlation coefficient obtained by the ordinary method.
The correlation coefficient of the ellipse shown in Fig. 1 is zero. The correlation coefficient of the random dataset in the ellipse is also close to zero at 0.013.

Though the ellipse and the rotated ellipse have the same major and minor radii, different results are obtained. I will explain this fact in the next articles.

On the other hand, the following equation is given from Eq.(14).

$$\frac{b}{a}=\sqrt{\frac{1-R_{XY}}{1+R_{XY}}} \qquad (15)$$

The *b/a* value is in the range of 0 to 1.
For example, substituting the correlation coefficient of the random data set 0.873 into Eq.(15), we get *b/a*＝0.260, which closely matches the theoretical value of the ellipse (*b/a*＝0.268).

Thus, the *b/a* value indicates the fluctuation range of the ellipse in the y-axis direction, and represents the correlation more quantitatively.

References

1. Summary of trigonometric identities
http://www2.clarku.edu/faculty/djoyce/trig/identities.html

Postscript

If you have any questions or comments regarding this article, please email to:
k.n.-zue06114@outlook.jp

If you can write Japanese, please write in Japanese.