Geology 659 - Quantitative Methods

Sampling, Autocorrelation and Crosscorrelation

tom.h.wilson
wilson@geo.wvu.edu

Department of Geology and Geography
West Virginia University
Morgantown, WV
We live in a digital world. Digital data is ubiquitous whether in the form of digital elevation data, stream gauge data ... or that DVD of your favorite movie. The choice of the sample interval is critical to accurate representation of information contained in continuous or analog data.

How do we decide what the sample rate should be? How often should we sample rainfall, ion concentrations, surface topography, etc., to accurately represent it?

The example of a rotating wheel with spokes serves as interesting example of the possible influence of sample rate on the accuracy of conclusions drawn for digital data sets.
Start with a wheel that rotates at constant frequency and then sample it at varying rates.

Assume the period of rotation ($\tau$) is 1 second and that we sample the motion of the wheel every 1/4 second.

In this case $\tau_{\text{observed}} = \tau_{\text{actual}}$
Let $\Delta t = 1/2$ second

We estimate $\tau = 1$ but we are unable to determine the direction of wheel rotation.
Let $\Delta t = \frac{3}{4}$ second

Our estimate of $\tau = 3$ seconds is in error - by 300%
When $\Delta t = 1$ second, the wheel appears to remain at the starting point but actually makes a single rotation between samples.
In the **second case**, we deal with situations where the sample interval $\Delta t$ is constant and the period of rotation $\tau$ varies. Take the case where $\Delta t = 1/2$ second and $\tau = 4$ seconds.

We have no problem in this case accurately delineating the period of rotation.

**Note** that $f = 1/\tau = 1/4$ second, and that $f \Delta t = 1/8$th. $f \Delta t$ is the fraction of a cycle the wheel has moved in time $\Delta t$. 
When $\Delta t$ equals $1/2$ we also correctly identify rotation periods of 1.5 seconds.
As before, when $\tau = 1$, and $\Delta t = 1/2$, we are still able to identify the period as 1 second.
However, when $\tau = 3/4$ sec, the frequency $f = 1/\tau = 1.33$ cycles per second. Between consecutive samples, the wheel will turn through $f\Delta t$ cycles (check units). $f\Delta t = 2/3$rd cycles between samples. The output period is 1.5 seconds and the output frequency is $2/3$rd cycle per second.

Our estimate of $\tau$ in this case is twice its actual value.
When $\tau = 2/3$ seconds, the frequency of rotation, $f=1.5$ cycles/sec. Thus the wheel turns through $f\Delta t$ or $3/4$ cycles between samples. The wheel appears to move counterclockwise with frequency equal to $1/2$ cycle per second.

When $\tau = 1/2$ second, the frequency of rotation, $f=2$ cycles/sec. $f \Delta t =1$. Thus the wheel turns through one cycle between samples. The wheel appears fixed. The output frequency is 0.
A plot of input versus output (sampled) frequency reveals saw tooth like variations of output frequency as input frequency increases. The output frequency never increases beyond 1 cycle per second in the case where motion is sampled every 1/2 second.
The Nyquist frequency is the highest frequency one could observe using a given sample rate.

\[ f_N = \frac{1}{2\Delta t} \]
Poor choice of sample will transform higher input frequencies into lower output frequencies. This phenomena is referred to as frequency- or signal-aliasing.
Autocorrelation and Crosscorrelation

Discussion of basic concept around the following figure from Davis’s text

![Graph](image-url)
Recall the basic definition of the correlation coefficient:

$$r_{xy} = \frac{\text{covar}_{xy}}{s_x s_y}$$

and also recall the basic definitions of the covariance and standard deviation.

$$\text{cov}_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

$$s_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$
Combine these terms, assume 0-average, and consider how $r$ will be simplified.

You should get:

$$r_{xy} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

Explicit reference to summation elements $i$ through $n$ has been left out for simplicity.
Consider the following sequence of numbers. Note that the set of numbers has 0 average.

\[1 \frac{1}{2}, -1, -\frac{1}{4}\]

Verify that \( r \), the correlation of the series with itself, equals 1.
Computational steps of the autocorrelation function are illustrated graphically below.

\[ a_0 = 3.5 \]

\[ 2.25 + 1 + \frac{1}{4} = 3.5 \]
Autocorrelation involves the repeated computation of the correlation coefficient $r$ between a series and a shifted version of the series. The shift is referred to as the lag. The computation of the autocorrelation for our simple function with lag = 1 is shown below.
The lag 2 value of the autocorrelation is computed in the same way, but after shifting an image of the input series two sample values relative to the original sequence.
The resultant autocorrelation function consists of 3 terms.

To convert these numbers into correlation coefficients we need only normalize each term in the series by 3.5.
We’ll consider the mathematical representation of the autocorrelation function leading to

\[ a_\tau = \sum_{t=0}^{n-1} f_t f_{t+\tau} \]

In its discrete form, and

\[ a(\tau) = \int_{-\infty}^{\infty} f(t)f(t+\tau) \, dt \]

In its continuous form.
**Autocorrelation**

Let’s take another look at this diagram from Davis and see if we understand it a little better.
**crosscorrelation**

Take the following two series of numbers, assume they are paired observations and compute the correlation coefficient between them.

Given series 1: 2, -1, -1, and

series 2: 1.5, -1, -0.5

Note that both series have 0 mean value.
\[ r_{xy} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \]

\[ r_{xy} = \frac{3+1+0.5}{\sqrt{(4+1+1)(2.25+1+0.5)}} = \frac{4.5}{4.58} = 0.983 \]

X column -> X
Y column -> Y
Number of used data points: 3

\[ a = 0.7500000 \]
\[ b = 0 \]
\[ \text{Sum Sqr} = 0.1250000 \]
\[ \text{StdDev} = 0.3535339 \]

Covariance Matrix
\[ \text{cvm}[1,1] = 0.16666667 \]
\[ \text{cvm}[1,2] = 0 \]
\[ \text{cvm}[2,2] = 0.33333333 \]

Goodness of Fit Statistics ...
\[ R^2 = 0.96428571 \]
\[ \text{Correlation} = 0.98198051 \]

Parameter Statistics
\[ \text{Parameter a:} 0.7500000 \]
\[ \text{StdDev:} 0.14433757 \]

\[ \text{Parameter b:} 0 \]
\[ \text{StdDev:} 0.20412415 \]
Thursday we’ll meet in lab.

Continue your reading of Chapter 4.