Data analysis of single time series

Eli Galanti Math-Meth 2018

December 3, 2018

Motivating examples: Tropical Pacific SST

Show Matlab example

Mean, variance and STD

A time series x_i could be any quantity as function of time or space.
The mean is

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

the variance is

$$\overline{x'^2} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2$$

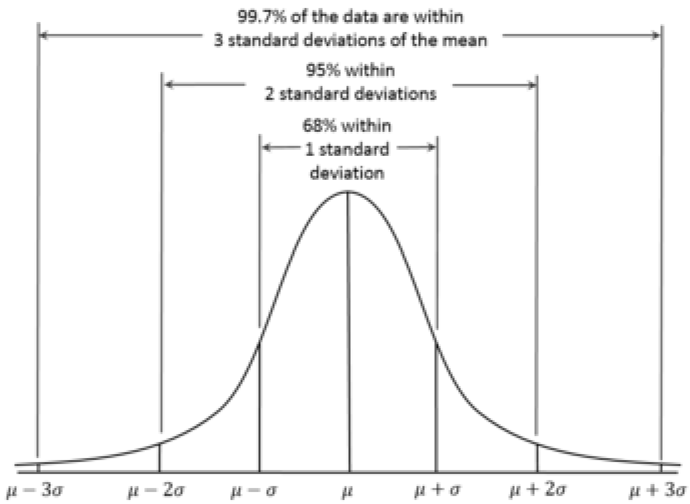
note that the normalization is sometimes taken as $N-1$.
The standard deviation σ is the square root of the variance

$$\sigma = \sqrt{\overline{x'^2}}$$

If the data has a normal distribution

$$f(x,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\overline{x})^2}{2\sigma^2}\right]$$

then the STD reflects the percentage of data being inside.



The mean and variance can be also applied to a collection of data points with no relation between neighboring points.

Auto-covariance and auto-correlation

The auto covariance function is

$$\phi(L) = \frac{1}{N-L}\sum_{k=1}^{N-L} x'_k x'_{k+L} = \overline{x'_k x'_{k+L}};\ L = 1, 2, 3, ...$$

where the prime denotes a deviation from the mean. The auto-correlation is then

$$r(L) = \frac{\phi(L)}{\phi(0)}$$

Show Matlab example

Composites

Another way to identify a pattern is to isolate regions in the time series that have something in common and plot them together. For example, identify local maximum and plot around that maximum.
Discuss Matlab 'findpeaks'

Show Matlab example

Regression

Fitting data to a function with unknown coefficients. Given a set of observations (x_i, y_i) for $i = 1 : N$, we can, for example, fit a linear function

$$\hat{y} = a_0 + a_1 x$$

by demanding that the measure

$$Q = \frac{1}{N}\sum_{i=1}^{N}\epsilon_i^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y} - y_i\right)^2$$

is minimized, so that

$$\frac{dQ}{da_0} = 2a_0N + 2a_1\sum x_i - 2\sum y_i = 0$$

$$\frac{dQ}{da_1} = 2a_0\sum x_i + 2a_1\sum x_i^2 - 2\sum x_iy_i = 0$$

and the solutions for a_0, a_1 are

$$a_1 = \frac{\overline{xy} - \overline{x}\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\overline{x'y'}}{\overline{x'^2}}$$

$$a_0 = \overline{y} - a_1\overline{x}$$

We will discuss this much more when talking about combining data with models since the function $\hat{y}$ is in fact a model.

Spectral analysis

There are many ways to decompose the time variability. We will look at the Discrete Fourier Transform (DFT) and use the Matlab FFT. Given an evenly spaced time series $y_n, n = o : N-1$, the DFT coefficients are

$$Y_k = \sum_{n=0}^{N-1} y_n e^{-i2\pi kn/N}$$

where y_n can be complex. The time series can be then reconstructed using

$$y_n = \frac{1}{N}\sum_{k=0}^{N-1} Y_k e^{i2\pi kn/N}$$

since $e^{-i\theta} = \cos\theta - i\sin\theta$, each Fourier coefficient Y_k is a measure for the projection of the time series on a sine and cosine functions. Remember that

$$\int_0^{\pi}\sin(nx)\sin(mx) = \begin{cases}\pi/2 & n = m\\ 0 & n \neq m\end{cases}$$

so that in addition to the orthogonality consequences, we require that the data is defined in multiples of half the wave length. Some things to keep in mind:

- The reason we need both functions is because the data might be in a specific phase compared to these functions.
- Y_0 is simply the sum of y_n (almost the mean...).
- we assume the boundaries are cyclic.

Some useful quantities: if Δt is the sampling interval, the sampling rate is

$$f_s = 1/\Delta t$$

, N is the number of samples, and T is the time length, then

$$\Delta f = 1/(\Delta t N) = 1/T$$

is the frequency interval in Hertz, and the maximal frequency resolved is $f_{max} = N/T = f_s$.
Because we need at least 2 data points to resolve a sine function, the maximum resolved frequency is

$$f_{max} = f_s/2$$

(called the Nyquist frequency). For this frequency, the phase cannot be determined so that $imag(y_{N/2}) = 0$.

Show Matlab example
Show ENSO fft Matlab example

Spectral analysis in space

Can use DFT if boundaries are cyclic. On a sphere, can use Legendre Polynomials for the polar direction, sine and cosine for the azimuthal direction, and spherical Bessel functions for the radial direction.