Time Series

Lesson 1

Grant Foster

Change is the essential property of the universe – Spock

The mathematical discipline which focuses on analyzing and understanding changes over time is *time series analysis*.

What is Time Series?

There are **two senses** of the phrase:

1: A *process* by which data change over time,
i.e., the “rules” by which things change.
Examples: flip a coin – free fall
May not have happened yet

2: A set of data at a set of times.
Also known as a *realization* of a time series process.
Might not (probably don’t) know the process

Deterministic .vs. Stochastic

- **Deterministic**: execute the TS process many times, get the same result every time.

- **Stochastic**: execute the TS process many times, get different results.

Two numbers for each "observation":

1: The *data* (usually call it x)

2: The *time* (usually call it t)

Therefore a TS in the sense of a *realization* is a set of data pairs (t_j, x_j), for $j = 1, 2, ..., N$ values, giving the time t_j and value x_j of the j^{th} observation.

When the times are evenly spaced (referred to as *evenly sampled* time series), it's not uncommon to omit recording the times. Instead we either record the index numbers of the data values (from 1 to however many we have), or we may not even do that, simply assuming that they're "understood."

This is a common practice for many treatments of time series, often thought of as a set of single numbers x_j rather than number pairs. That's fine **if** the times are evenly spaced, but in the very common case of uneven sampling it just won't do.

Time is not necessarily “time”

Coin-flip example: recorded time as the flip number. Time starts at 1, goes up to the number of flips we execute, and takes only integer values.

By the nature of the process, time is a *discrete* variable for the coin flip. A considerable part of the theory of time series involves processes for discrete time series.

Time is a continuous variable

Of course we can't analyze a continuum of data! In fact any *observed* time series consists of a finite number of data pairs. Even when the time variable is continuous, if the observations are taken at regular intervals of time we can apply many of the methods used for discrete time series.

But if the times of observations are irregularly spaced (referred to as *uneven sampling* or *irregular sampling*) a number of complications arise in the analysis.

Time is Special

- It's an *independent* variable, truly and completely.
- There's an "arrow of time" – it goes from past to future.
- Time measurements usually treated as error-free.

Goals

In a broader sense, *understand* the time series process.

TS process: understand how the data will behave if we execute it.

TS realization: identify the *process* which generates it.
Which part of the process is deterministic, creating the *signal*?
Which part is stochastic, creating the *noise*?
Understand the nature of each.

End up with a useful *model* of the process.

Another Goal

Predict the future course of the time series variable.

Even if we don't necessarily know the process with confidence, we may be able to approximate it sufficiently to make useful (if uncertain) predictions.

Applies to both deterministic and random time series, and to combinations of the two; even if the process is entirely random, it may be complicated enough that we can calculate useful predictions far enough into the future to be of palpable benefit.

Notation

Most statistics uses $\mathbf{E}(x)$ to denote the *expected value* of a random variable x. I prefer to use $\langle x \rangle$.

Angle brackets around *random variable* $\Rightarrow$ expected value

$$\langle Q_n \rangle = \text{expected value of } Q_n,$$

Around *observed data* $\Rightarrow$ *average* value

$$\langle Q \rangle = \frac{1}{N} \sum_{n=1}^{N} Q_n.$$

Abstract Index Formalism

Suppose a TS is the sum of a deterministic part and a stochastic part:

$$x_n = f(t_n) + \varepsilon_n$$

The "n" subscript on x in this equation refers to a specific, single value of the observed variable x at a specific, single value of the time t. If this holds for all n, then we have a separate copy of this equation that holds for every index value n.

Abstract Index Formalism (cont.)

Sometimes it's desirable to denote, rather than a single value of the observable, *all* the values taken together as a whole.

There's considerable advantage in treating the time series data as a *vector*, in which case it's even more advantageous to treat all the individual quantities as a unified whole rather than a disparate set of numbers.

Abstract Index Formalism (cont.)

Convention (not always followed!) a subscript taken from the lowercase version of the latin alphabet, like n in the equation above, will denote an individual value of the times series, as it usually does. If that equation holds for the entire time series, then we have multiple copies of that equation, one for each index value.

Abstract Index Formalism (cont.)

Convention (not always followed!) a subscript taken from the lowercase version of the latin alphabet, like n in the equation above, will denote an individual value of the times series, as it usually does. If that equation holds for the entire time series, then we have multiple copies of that equation, one for each index value.

A subscript taken from the Greek alphabet will denote, not an individual member of the time series, but the *vector* of all values.

Abstract Index Formalism (cont.)

The equation

$$x_\alpha = f_\alpha + \varepsilon_\alpha,$$

indicates that the *vector* x_α is decomposed into the deterministic and stochastic *vectors* f_α and ε_α.

It follows that equation

$$x_n = f_n + \varepsilon_n,$$

must hold true for each individual component of those vectors, i.e., for each value of n.

Abstract Index Formalism (cont.)

Second equation holds for each time index value, but first equation is not a statement about individual time index values, rather it's an equality between two vectors.

MANY notations for vectors

Statistics: set a symbol in bold-face type

$$\mathbf{x} = \mathbf{f} + \mathbf{e}.$$

Physics: place an arrow above a symbol

$$\vec{x} = \vec{f} + \vec{\varepsilon}.$$

Quantum mechanics: place a vertical bar to the left, angle bracket to the right

$$|x\rangle = |f\rangle + |\varepsilon\rangle.$$

MANY notations for vectors (cont.)

They're all the same equation, just using different notation.

Keep in mind that in the abstract-index notation, the abstract marker is *not* an index, it's just a marker indicating the vector nature of the given quantity.

Einstein Summation Convention

We often have to sum an expression over all values of the time index, or all values of multiple time indexes. Example: the sum of the products of the values of two time series x and y is

$$\sum_{n=1}^{N} x_n y_n.$$

Einstein Summation Convention (cont.)

Useful convention: when using Greek indices to denote vectors, any index which is *repeated* indicates the *dot product* (or *inner product* or *transvection*) of the vectors, which is computationally equivalent to the *sum over all index values*. Known as the *Einstein summation convention.*

Hence the same quantity indicated before can simply be written

$$x_\alpha y_\alpha.$$

Einstein Summation Convention (cont.)

Another notation for *inner product*:

$$
\begin{aligned}
&\mathbf{x}^{\mathrm{T}}\mathbf{y} \\
&= \vec{x} \cdot \vec{y} \\
&= \langle x | y \rangle \\
&= \sum_{j=1}^{N} x_j y_j \\
&= x_\alpha y_\alpha .
\end{aligned}
$$

Einstein Summation Convention (cont.)

Eliminates the need to write repeated summation symbols. After you become familiar with it, will actually help you be clear which variables are summed, and in which way. We can extend the convention to multiple summations using multiple indices, so that for example

$$\sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{n=1}^{N} M_{jk} c_n x_j y_k t_n = M_{\alpha\beta} c_\lambda x_\alpha y_\beta t_\lambda.$$

Notation

In private calculations, use the notation that works for you.
In publications, use notation that's clear to your readers.

Best notational advice I ever got:

Notation should be your servant,
not your master.

Visual Inspection

Eye-brain combination is one of the most potent pattern-recognition systems known. A first step in time series analysis (alas, one of the too-often neglected steps) is to

graph the data and look at it.

Process is referred to as *visual inspection*.

Never underestimate its power.

Visual Inspection (cont.)

Risk of visual inspection: Eye-brain combination is all too easily fooled, far too likely to see patterns or processes which are really just noise – "pictures in the clouds" that aren't real. Don't neglect visual inspection, it's too potent to ignore, but do be skeptical about what you see might be there.

Decomposition of a Time Series

Time series can be completely random (coin-flip series)

Can be completely deterministic (path of a spacecraft through empty space)

More often, time series are a combination of both, having a deterministic part and a random part.

Deterministic part sometimes called *signal*

Random part also called *stochastic* part, sometimes called *noise.*

One type of noise: *measurement error*

Difference between any physical process and our *measurement* of that process.

Example: the brightness of a star can be measured, usually on a scale called *magnitude*; if the star is perfectly constant then its magnitude will be constant. When we measure the brightness, our measurments are bound to be imperfect. Difference between the brightness and the *measured* or *estimated* brightness is called the *measurement error*.

Random *Process*

Not all randomness is measurement error.

The *process itself* can include randomness, e.g. coin flip.

Processes can be *effectively* random, e.g. rainfall amount.

Decomposition of Time Series

• Separate into *deterministic* and *stochastic* parts (signal and noise).

• Most common way: *additive* model.

$$\text{data} = \text{signal} + \text{noise}$$

$$x(t) = f(t) + \varepsilon(t)$$

Further Decomposition

Separate signal into periodic part (*cycles* or *seasonality*) and non-periodic part (*trend*)

Additive model:

$$f(t) = T(t) + S(t).$$

Not Necessarily Additive

Perhaps $f(t) = T(t)S(t)$ (multiplicative).

Or maybe $f(t) = T(t)^{S(t)}$ (exponential).

Generally, $f(t) = F(T(t), S(t))$ (any functional relationship).

Only limit is your imagination!

But – additive model is most common.

Mix and Match

Example: signal is multiplicative combination of trend and seasonality, time series is additive model of signal and noise:

$$x(t) = T(t)S(t) + \varepsilon(t).$$

Most generic:

$$x(t) = F(T, S, \varepsilon).$$

General principle: simplicity.

Stationary Time Series

Stochastic $\Rightarrow$ the most we can know is the *probability distribution* of values.

Each value x_j may follow a different distribution

$$\text{Prob.} = P_j(x_j)\ dx_j.$$

Of course

$$P_j(x_j) \geq 0,$$

$$\int P_j(x_j)\ dx_j = 1.$$

Relationship among different values

Even if completely random, different values may be related.

Example: if one value is large, the next may be more likely to be large also (or perhaps more likely to be small).

Example: the value 12 time "steps" later may be more likely large or small.

Only limit to relationships is your imagination.

Relationships (cont.)

Express relationships between different values through the *joint probability distribution*

$$\text{Prob.} = f(x_{n_1}, x_{n_2}, ..., x_{n_j})\ dx_{n_1}\ dx_{n_2}\ ...\ dx_{n_j}.$$

Purely Deterministic

If there's no randomness, we can be certain what the value will be. We might know, e.g. that $x = 0$.

Then the *probability distribution* expresses certainty. The pdf is the **Dirac delta function** $\delta(x)$.

Zero for $x \neq 0$ (so the probability $x \neq 0$ is zero, i.e. impossible).

It's "infinite" for $x = 0$, but "just the right size" infinity that

$$\int \delta(x)\ dx = 1.$$

Dirac Delta (cont.)

Most assuredly **not** a function (topic for another course).

Version for discrete (rather than continuous)

$$\delta(x) = \begin{cases} 1 & x = 0 \\ \\ 0 & \text{else} \end{cases}$$

Discrete version *is* a function (with limited domain).

Stationary Time Series

pdf is the same for all values

$$P_{n_1}(x_{n1}) = P_{n_2}(x_{n_2}) \quad \text{for all } n_1,\ n_2.$$

Can also write as

$$P_n(x_n) = P_{n_k}(x_{n+k}) \quad \text{for all } n,\ k.$$

k is the *lag* between the values.

Stationary Time Series (cont.)

Essence: pdf is *time-translation invariant*,
i.e. translating through time doesn't change the behavior.

Stationary Time Series (cont.)

How about dependencies between different values?

Joint pdf is also time-translation invariant if, for any choices $x_{n_1}, x_{n_2}, ..., x_{n_j}$, the joint probability distribution is unchanged when all the relevant times $t_{n_1}, t_{n_2}, ..., t_{n_j}$ are offset by the same constant τ, i.e.,

$$P(x(t_{n_1}), x(t_{n_2}), ..., x(t_{n_j}))$$

$$= P(x(t_{n_1} + \tau), x(t_{n_2} + \tau), ..., x(t_{n_j} + \tau)).$$

Stationary Time Series (cont.)

If (and *only* if) the times are equally spaced, then a constant offset in time by τ corresponds to a constant offset of the *index* value by k, and we can state the requirement for time-translation invariance as

$$P(x_{n_1}, x_{n_2}, ..., x_{n_j}) = P(x_{n_1+k}, x_{n_2+k}, ..., x_{n_j+k}).$$

Stationary Time Series (cont.)

Definition: a time series is *strongly stationary* $\Leftrightarrow$ all pdfs for individual values and joint pdfs for multiple values are time-translation invariant.

Essentially, it means that the essential nature doesn't change.

Values can still be related to other values (to "conditions at the time"), but if we recreate the same conditions, we'll get the same pdf.

Stationary Time Series (cont.)

Strongly stationary is extremely stringent, and extremely hard to demonstrate for observed time series (it's a condition on *all* joint pdfs). Almost always use a weaker (but useful) condition.

Stationary Time Series (cont.)

Require only that the first two *moments* of probability be time-translation invariant. Hence expected value (1st moment) must be the same for all times

$$\langle x_j \rangle = \langle x_{j+k} \rangle,$$

for all index values j and offsets k.

Stationary Time Series (cont.)

Likewise for the expected product of two values (2nd moments)

$$\langle x(t_1)x(t_2)\rangle = \langle x(t_1+\tau)x(t_2+\tau)\rangle,$$

or all times t_1 and t_2, and all time offsets τ. If the times are evenly spaced, we can equivalently say

$$\langle x_j x_k\rangle = \langle x_{j+h} x_{k+h}\rangle,$$

for any index values j and k, index offset h.

Stationary Time Series (cont.)

Such a time series is said to be *weakly stationary*. Unless otherwise stated, the word "stationary" means weakly stationary.

Much less stringent than strongly stationary, and much more practical.

Stationary Time Series (cont.)

Why so useful? Because we so often estimate things based on many data values (parameters of a model, summary statistics, ...).

When we do, the *central limit theorem* ensures those estimates follow the normal distribution, which is completely characterized by its mean and standard deviation. The first two moments of the TS often enable us to compute the first two moments of the parameters/etc., hence its normal distribution.

White Noise

One particular type of random process is extremely important in time series analysis: *white noise*. It's defined by the fact that it is stationary, and that any two different values have no *covariance*.

Covariance

The *covariance* of two variables is the expected value of the product of their deviations from their respective means, i.e.

$$cov(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle.$$

Covariance (cont.)

We can use the fact that the expected value is not a random variable (rather it's a property of a probability distribution)

$$cov(X, Y) = \langle XY \rangle - \langle X \rangle \langle Y \rangle - \langle X \rangle \langle Y \rangle + \langle X \rangle \langle Y \rangle$$

$$= \langle XY \rangle - \langle X \rangle \langle Y \rangle.$$

Therefore the covariance of two quantities is the difference between the expected value of their product, and the product of their expected values.

White Noise (cont.)

Since white noise is stationary, its expected value doesn't depend on time

$$\langle x_n \rangle = constant = \mu,$$

where μ doesn't depend on time, so it doesn't depend on the time index n. The 2nd moment of the distribution for x_n must also be constant through time, so that

$$\langle x_n^2 \rangle = constant = \mu^2 + \sigma^2,$$

which defines σ^2, the *variance* of the time series values.

White Noise (cont.)

The variance is, of course, just the covariance of a value with itself

$$\sigma^2 = \langle x_n^2 \rangle - \langle x_n \rangle^2.$$

White Noise (cont.)

The fact that different values have no covariance means

$$\langle x_j x_k \rangle - \langle x_j \rangle \langle x_k \rangle = 0,$$

for $j \neq k$. We can therefore say in general that for white noise

$$\langle x_j x_k \rangle = \mu^2 + \sigma^2 \delta(j - k),$$

where $\delta(n)$ is the discrete Dirac δ-function. A time series process obeying these requirements is *white noise*.

i.i.d. Noise

For i.i.d noise, different values are actually *independent*, meaning the joint pdf for multiple values is just the product of the one-variable pdf for the different values

$$P(x_a, x_b, ..., x_j) = P_a(x_a)P_b(x_b)...P_j(x_j).$$

Furthermore, since this type of noise is *identically distributed* all the functions $P_a(x_a), P_b(x_b), ...$ are the *same* function.

i.i.d. Noise (cont.)

It's not too hard to see that for i.i.d. noise, all the joint probability distributions are time-translation invariant. Hence i.i.d. noise isn't just stationary, it's *strongly* stationary.

Zero-Mean Gaussian i.i.d. Noise

Gaussian noise follows the normal probability distribution

$$P(x) = \frac{e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}}{\sigma\sqrt{2\pi}}.$$

It often happens that the mean value μ is zero, giving *zero-mean Gaussian i.i.d. noise*

$$P(x) = \frac{e^{-\frac{1}{2}x^2/\sigma^2}}{\sigma\sqrt{2\pi}}.$$

Correlation

Covariance of two quantities is the difference between the expected value of their product, and the product of their expected values.

<u>*Correlation*</u> between two quantities is their covariance, *scaled* by the square root of their variances (i.e., by their standard deviations).

Correlation (cont.)

Variance is covariance of a variable with itself, so

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{cov(X,X)cov(Y,Y)}}.$$

Correlation (cont.)

If we rescale the quantities by multiplying them by constants α and β, defining new quantities

$$\tilde{X} = \alpha X, \quad \tilde{Y} = \beta Y,$$

then the correlation of the re-scaled quantities is the same as that of the original quantities

$$corr(\tilde{X}, \tilde{Y}) = corr(X, Y).$$

Correlation (cont.)

It's easy to see that the correlation of a quantity with itself is exactly equal to 1. The correlation of any two quantities is always between -1 and +1.

Problem 1

This is for general online discussion; feel free to post your answer in the comment section.

- Give an example of a time series process which is white noise, but is not i.i.d.

Problem 2

This is for general online discussion; feel free to post your answer in the comment section.

- Give an example of a time series process which is weakly stationary but not strongly stationary

Problem 3

We said that if random variables x and y are transformed by re-scaling, so the new varibles are $\tilde{x} = \alpha x$ and $\tilde{y} = \beta y$ (α and β constants), then the new variables have the same correlation as the original variables.

Suppose we apply an affine transformation, so the new variables are $\tilde{x} = \alpha_1 x + \alpha_2$ and $\tilde{y} = \beta_1 y + \beta_2$ (with $\alpha_1,\ \alpha_2,\ \beta_1\ , \beta_2$ constants). Show that the new variables have the same correlation as the old.

Problem 4

I suspect this is a tough one; don't worry too much about it (or, maybe it's easy?).

Show that the correlation between any two random variables x and y cannot be greater than 1.