NONLINEAR TIME SERIES ANALYSIS, WITH APPLICATIONS TO MEDICINE

José María Amigó

Centro de Investigación Operativa, Universidad Miguel Hernández, Elche (Spain)

LECTURE 4

NONLINEAR METHODS IN MEDICINE I: GENERAL TECHNIQUES

OUTLINE

1. **What is nonlinear time series analysis**
2. **Practical issues**
3. **Conventional linear methods**
4. **State-space reconstruction**
5. **Noise reduction**
6. **Testing stationarity & determinism**
7. **Surrogate data testing**
8. **References**

1. What is nonlinear TSA?

- Chapter 2: From first principles to phenomenology (*bottom up approach*)
- This chapter: From phenomenology to the system (*top down approach*)

1. What is nonlinear TSA?

- **Working hypothesis:** Nature is deterministic, nonlinear, and dissipative.
- **Picture:** Data are measurements at successive times of a system evolving usually in continuous time,

$$\boxed{t \mapsto x(t) \in \Omega} \quad ===> \quad s_n = s(x(t_n))$$

  where $s : \Omega \to \mathbb{R}$ is called the *measurement function* and t_n the *measurement times*.
- **Scope:** Extract information from the data
- **But** there are quite a few practical issues.

1. What is nonlinear TSA?

Examples of time series.

- Meteorological observations (temperature, pressure, humidity, ...)
- Financial data (trading price, market averages, exchange rates, ...)
- Biomedical records (EEG, ECG, epidemiological data, ...)

1. What is nonlinear TSA?

Procedure.

1. Reconstruction of the state space
2. Noise reduction
3. Test stationarity and determinism
4. Characterize the reconstructed attractor
5. Check the validity of the conclusions (surrogate data)

Remark. *The exact order depends on whether the methods need a reconstructed state space.*

2. Practical issues

Time series analysis involves a number of practical issues.

1) Sampling and delay time.

In practice, data are sampled at a constant *sampling frequency* $1/\Delta t$, i.e.,

$$s_0, s_1, ..., s_n, ... = s(x(0)), s(x(\Delta t)), ..., s(x(n\Delta t)), ...$$

where Δt is the *sampling time*, and $n\Delta t \equiv t_n$ are the measurement times.

Δt *should be much smaller than the variation scale of* $x(t)$.

A better option:

$$s_0, s_\tau, ..., s_{n\tau}, ... = s(x(0)), s(x(\tau\Delta t)), ..., s_{n\tau} = s(\mathbf{x}(n\tau\Delta t)), ...$$

The parameter $\tau \geq 1$ is called the *lag* or *delay time*.

2. Practical issues

2) Finiteness.

Real-world time series are finite –but most quantities (entropies, λ, dimensions, etc.) involve infinite limits!

Remedies: Produce as many data as possible, use alternative estimators, use finite-size correction terms, ...

General rule: It suffices to obtain the *scaling behavior*.

2. Practical issues

3) Data contamination

Any error in the determination of the states.

- If the error is propagated by the dynamic: *dynamical or multiplicative noise*
- If the error is not propagated by the dynamic: *observational or additive noise*

Modelling additive noise.

$$s_n = s(x(n\Delta t)) + w_n$$

where $(w_n)_{n\geq 0}$ is *white noise* (i.e., an i.i.d. random process).

One can use *noise-reduction techniques*.

2. Practical issues

4) Stationarity.

Stationarity means hat the statistical properties (averages, deviations, etc.) do not depend on which part of the time series I am considering.

Causes of nonstationarity.

- If the system is random: the process is nonstationary.
- If the system is deterministic: transients, change of parameters.
- Time series too short to capture the longest characteristic time scale.

3. Conventional linear methods

Conventional linear methods are useful for some basic purposes, like checking stationarity or discriminating randomness from chaoticity.

1) Stationarity with 1st and 2nd order statistics

If $\mathbf{x} = x_1, ..., x_N$ is a nonstationary time series, then its *mean*

$$\langle \mathbf{x} \rangle = \frac{1}{N} \sum_{n=1}^{N} x_n,$$

and/or the *standard deviation*

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (x_n - \langle \mathbf{x} \rangle)^2}$$

change with N.

3. Conventional linear methods

Test.

1. Calculate $\langle \mathbf{x} \rangle$ and σ for the 1st half $(0 \leq n \leq \lfloor N/2 \rfloor)$
2. Calculate $\langle \mathbf{x} \rangle$ and σ for the 2nd half $(\lfloor N/2 \rfloor + 1 \leq n \leq N)$
3. If the means of the two halves differ by more than a few *standard errors* for each half $(\sigma/\sqrt{N/2})$, then stationarity is a problem.

Other possibility: Plot a *histogram* of the probability distribution $p(\mathbf{x})$ (number of bins $\sim \sqrt{N}$). $p(\mathbf{x})$ should remain approximately constant for different $\mathbf{x}$'s if the data source is stationary.

2) Determinism with Fourier analysis

Assumption: $\mathbf{x} = x_1, ..., x_N$ can be represented by a superposition of sines and cosines of various amplitudes and frequencies.

- *Fundamental frequency*: $f_0 = 1/N$
- *Highest frequency*: $f_c = 1/2$ (*Nyquist frequency*)
- *Frequencies of the harmonics*: $f_\nu = \nu/N = \nu f_0$ $(1 \leq \nu \leq N/2)$

$$x_n \simeq \frac{a_0}{2} + \sum_{\nu=1}^{N/2} \left(a_\nu \cos\frac{2\pi\nu n}{N} + b_\nu \sin\frac{2\pi\nu n}{N} \right)$$

- *Amplitudes:*

$$a_\nu = \frac{2}{N}\sum_{n=1}^{N} x_n \cos\frac{2\pi\nu n}{N},\ b_\nu = \frac{2}{N}\sum_{n=1}^{N} x_n \sin\frac{2\pi\nu n}{N}.$$

3. Conventional linear methods

Definition (*Power spectral density*)**.** The power $S(f)$ at frequency νf_0 is

$$S_\nu = a_\nu^2 + b_\nu^2.$$

- Quasiperiodic signals have a finite number of sharp spectral peaks.
- Chaotic signals have a continuous (or 'broadband') spectrum, perhaps with some embedded peaks.
- But random noise has also a continuous spectrum.
- If the power spectrum diverges as $f^{-\alpha}$, the time series must be considered nonstationary.

Power spectrum of a chaotic signal

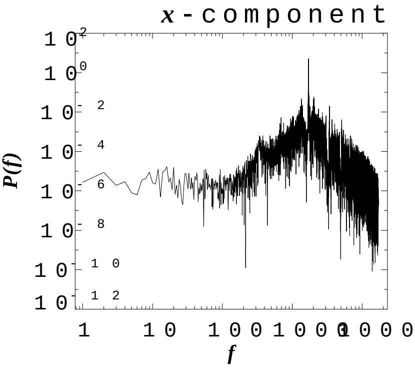


Figure. Power spectrum of a time series generated with the Rössler oscillator.

3. Conventional linear methods

Definition. The autocorrelation function of a stationary time series $\mathbf{x} = x_1, ..., x_N$,

$$G(k) \simeq \frac{\sum_{n=1}^{N-k}(x_n - \langle \mathbf{x} \rangle)(x_{n+k} - \langle \mathbf{x} \rangle)}{\sum_{n=1}^{N-k}(x_n - \langle \mathbf{x} \rangle)^2},$$

measures how strongly on average each data point is correlated with one k time steps away $(0 \leq k \leq N-1)$.

$G(0) = 1$
$G(k) = 1$ for perfect correlation
$G(k) = -1$ for perfect anticorrelation
$G(k) = 0$ for uncorrelated data
k is called the *lag*.

Facts.

- Random processes have decaying autocorrelations but the decay rate depends on the properties of the process.
- Autocorrelations of chaotic signals decay exponentially with increasing lag.

3. Conventional linear methods

Theorem (*Wiener-Khinchin*). The Fourier transform of $G(k)$ is the power spectrum.

$$S_\nu = G(0) + G(K)\cos\frac{2\pi\nu K}{N} + 2\sum_{k=1}^{K-1} G(k)\cos\frac{2\pi\nu k}{N},$$

where K is the maximum k, which should be taken about $N/4$.

4. State-space reconstruction

Assumption. $s_0, s_1, ..., s_n, ...$ has been obtained from a deterministic system evolving on an attractor:

$$s_n = s(x(n_0 + n\tau)\Delta t),$$

where

- $s : \Omega \to \mathbb{R}$ is the measurement function,
- $x(t)$ is the time evolution of the system
- Δt is the sampling time
- τ is the delay time
- n_0 allows for removing transients

Comments.

- If the system is deterministic, $s_{n+1} = \varphi(s_n, s_{n-1}, ...)$, but noise can blur this relation.
- The plots s_{n+1} vs $s_n, s_{n-1}, ...$ are called a *return maps*.
- The underlying dynamical system (Ω, f) is not known.
- If f is dissipative, dimension of the attractor $<$ dimension of Ω.

4. State-space reconstruction

The *state space reconstruction* allows to construct a new space Ω_{new} with an *equivalent* attractor, i.e.,

1. every point in Ω_{new} maps to a unique point by the dynamic,
2. the corresponding attractors in Ω and Ω_{new} have the same λ and dimensions.

The change of coordinates between Ω and Ω_{new} is called an *embedding*.

Theorem (Takens). Given a time series

$$s_{n_0}, s_{n_0+\tau}, s_{n_0+2\tau}, ..., s_{n_0+k\tau}, ...,$$

the *time-delay vectors*,

$$\mathbf{s}(n) = (s_{n-(m-1)\tau}, ..., s_{n-\tau}, s_n),$$

constitute an adequate embedding provided

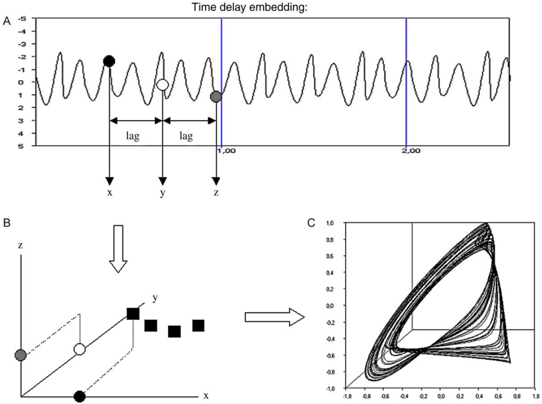
$$m \geq \lfloor 2D_0 \rfloor + 1.$$

where D_0 is the box-counting dimension of the attractor.

The parameter m is called the *embedding dimension* m.

4. State-space reconstruction

Illustration[1]



[1] C.J. Stam, Clin. Neurophys. 116 (2005) 226

4. State-space reconstruction

The state-space reconstruction has two parameters:

1. The embedding dimension m (all we know is $m \geq \lfloor 2D_0 \rfloor + 1$)
2. The delay time τ

How to choose them?

Choosing m: **False nearest neighbors**.

1. For each s_n find an s_l such that their distance in the reconstructed state space,

$$R_n(m) = \sqrt{(s_l - s_n)^2 + (s_{l-\tau} - s_{n-\tau})^2 + ... + (s_{l-(m-1)\tau} - s_{n-(m-1)\tau})^2}$$

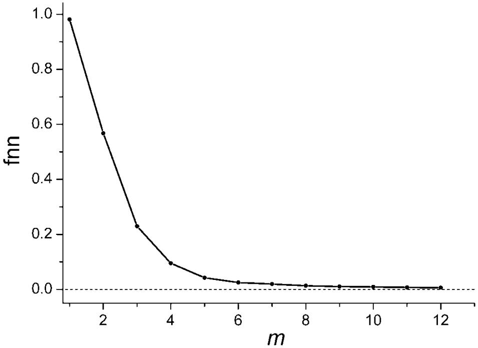
   is minimum. s_l is call the *nearest neighbor* of s_n.

2. Calculate $R_n(m+1)$. If $R_n(m+1) \gg R_n(m)$, then x_n and x_l are *false neighbors*. *Criterion for falseness*:

$$\frac{\left| s_{l-m\tau} - s_{n-m\tau} \right|}{R_n(m)} \gtrsim 15$$

3. Sample the time series and plot the fraction of *false nearest neighbors.*

Illustration[2].



[2] M. Perc, Eur. J. Phys. 26 (2005) 757

Choosing τ: Minimum mutual information.

- If τ is too small, the reconstructed attractor is stretched out along the diagonal of the embedding space.
- If τ is too large, the reconstructed attractor is excessively folded.

4. State-space reconstruction

Recommended method: Use the first minimum of the mutual information,

$$I(\tau) = \sum_{i=1}^{N} \sum_{j=1}^{N} p_{ij}(\tau) \log p_{ij}(\tau) - 2 \sum_{i=1}^{N} p_i \ln p_i,$$

- N is the number of bins in $[s_{\min}, s_{\max}]$
- p_i is the probability that x_n is in bin i
- $p_{ij}(\tau)$ is the probability that s_n is in bin i and $s_{n+\tau}$ is in bin j

The size of the bins is not critical as long as they are sufficiently small.

5. Noise reduction

There are noise reduction algorithms based on time-delay embedding.

Noisy measurements:

$$s_n = x_n + \zeta_n$$

where ζ_n is supposed to be a rv with fast decaying autocorrelation and no correlation with x_n.

$\sigma = \sqrt{\left\langle \zeta^2 \right\rangle}$ is called the *noise amplitude* or the *noise level*. σ can be estimated from a plot of the data or a correlation sum.

5. Noise reduction

Method *(nonlinear filtering)*. Let

$$\mathbf{s}(n) = (s_{n-(m-1)\tau}, ..., s_{n-\tau}, s_n)$$

be the embedding vectors.Then *replace the middle coordinate* $s_{n-\lfloor m/2 \rfloor \tau}$ *by the average middle coordinate of neighboring points*:

$$\hat{s}_{n+\lfloor m/2 \rfloor \tau} = \frac{1}{|B_\varepsilon(\mathbf{s}(n))|} \sum_{\mathbf{s}(k) \in B_\varepsilon(\mathbf{s}(n))} s_{n-\lfloor m/2 \rfloor},$$

where $B_\varepsilon(\mathbf{s}(n))$ is the m-dimensional ball of radius ε centered at $\mathbf{s}(n)$, and $|B_\varepsilon(\mathbf{s}(n))|$ is the number of points in it.

- Use $\varepsilon = 2\sigma$ or $\varepsilon = 3\sigma$.

Possible methods to test stationarity:

1. *Linear methods*.
2. *Nonlinear methods*: recurrence plots, cross prediction error statistic,...

Cross prediction error statistic will be explained below (as a test for determinsim)

6. Testing stationarity & determinism

Possible methods to test determinism:

1. Visual methods: Return maps, recurrence plots, ...
2. Fingerprints: Power spectrum, Lyapunov exponent,...
3. Forbidden ordinal patterns
4. *Kaplan-Glass test*
5. *Cross prediction error statistic*

6.1. Kaplan-Glass test.

Idea: Neighboring trajectories should point in about the same direction if the system is deterministic.

1. Quantize the *reconstructed* state space with a box partition.
2. Each pass of the trajectory generates a unit vector $\mathbf{e}_p$ determined by the entry and exit points in/from the box k.
3. The average directional vector of box k is

$$\mathbf{V}_k = \frac{1}{n}\sum_{p=1}^{n}\mathbf{e}$$

   where n is the number of total passes though the box k.
4. Let κ be the average length of all $\mathbf{V}_k/\left\|\mathbf{V}_k\right\|$: (i) $\kappa \approx 0$ random data; (ii) $\kappa \approx 1$ deterministic data.

6.2. Cross prediction error statistic.

Stationary deterministic systems are predictable, at least in the short term. Let $s_0, s_1, ..., s_N$ be a test data set in a longer series.

Question: $s_{N+1} =?$

Idea: Suppose $s_n = s(x_n)$, where $x_0, x_1, ..., x_N \in \mathbb{R}^d$ is deterministic

$$\left.\begin{array}{c} x_{n+1} = f(x_n), \\ x_k \simeq x_N \end{array}\right\} \Rightarrow x_{N+1} = f(x_N) \simeq f(x_k) = x_{k+1}$$

provided f is continuous.

Answer: $s_{N+1} \simeq s_{k+1}$

In practice though $x_0, x_1, ..., x_N$ are unknown. Use embedding vectors

$$\mathbf{s}(n) = (s_{n-(m-1)\tau}, ..., s_{n-\tau}, s_n)$$

to obtain states equivalent to the original ones. Thus:

$$\mathbf{s}(N+1) \simeq \mathbf{s}(k+1)$$

Even better:

$$\mathbf{s}(N+1) \simeq \frac{1}{|B_\varepsilon(\mathbf{s}(N))|} \sum_{\mathbf{s}(k)\in B_\varepsilon(\mathbf{s}(N))} \mathbf{s}(k+1)$$

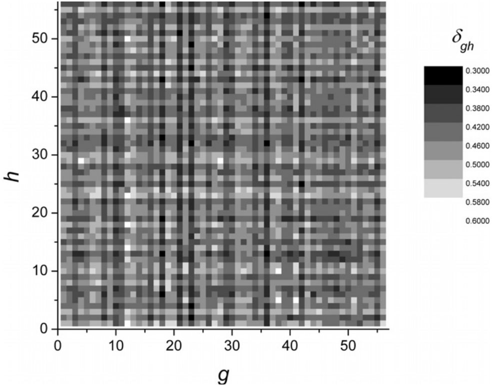
Prediction: The last component of $\mathbf{s}(N+1)$.

Determinism test:

1. Split the data $(s_n)_{n=1}^{N}$ into I short segments S_i $(N/I \approx 500)$
2. For all pair S_i, S_j make predictions in S_j using data of S_i.
3. Compute the rms of the predictions errors δ_{ij}
4. If $\delta_{ij} \gg$ average, then S_i is a bad model for S_j
5. Compare $(\delta_{ij})_{\min}, (\delta_{ij})_{\max}$ with the average

The matrix δ_{ij} is usually color-coded and plotted.

Illustration[3].



[3] M. Perc, Eur. J. Phys. 26 (2005) 757

7. Surrogate data testing

- Chaotic systems mimic (white and colored) noise.
- Random systems with nonuniform power spectrum can mimic chaos.

Recommendation. Test conclusions about whether a time series is chaotic with *surrogate data*[4].

Surrogate data are designed to mimic the statistical properties of chaotic data but with determinism removed.

[4] T. Schreiber and A. Schmitz, Physica A (2000) 346.

7. Surrogate data testing

1. Surrogate data with the *same probability distribution*: shuffle the data.
2. Surrogate data with the *same power spectrum*: use a surrogate series $Y = (y_n)_{n=1}^N$ with the same Fourier amplitudes but with random phases,

$$y_n = \frac{a_0}{2} + \sum_{\nu=1}^{N/2} \sqrt{S_\nu} \sin 2\pi \left(\frac{\nu n}{N} + r_\nu \right),$$

   where r_ν are $N/2$ uniform random numbers chosen from $0 \leq r_\nu < 1$.

Note. $p(Y)$ tends to be nearly Gaussian.

References

1. H. Kantz, and T. Schreiber, *Nonlinear time series analysis*. Cambridge University Press, 2004.
2. W.H. Press et al., *Numerical Recipes*. Cambridge University Press, 2007.
3. M. Small, Applied Nonlinear Time Series Analysis, World Scientific 2005.
4. J.C. Sprott, *Chaos and time series analysis*. Oxford University Press, 2003.