Nonparametric Time Series Analysis 1:

Measures of Complexity from Chaos Theory

Cees Diks

Center for Nonlinear Dynamics in Economics and Finance
University of Amsterdam

Pisa, St. Anna School, October 2006
Part I

Reconstructing dynamical systems from time series
1. Time series
   - Introduction

2. Nonlinear dynamics and chaos
   - Dynamical systems
   - Logistic map
   - Sensitive dependence
   - Hénon
   - Lorenz
   - Rössler

3. Attractor reconstruction from time series
   - Takens theorem
Time series

Sequence of measurements over time, for example
- Historic temperature records
- Electrical activity on the heart (ECG)
- Stock prices

Example: Atrial fibrillation (Koefib217)
Time series analysis

Main goals of time series analysis
- Modelling & prediction
- Characterisation & classification

Approaches
- Linear versus nonlinear
- Parametric versus nonparametric
Correlation integrals

- Originally used for estimation of dynamic invariants (deterministic dynamics)
- Measure of complexity (Hoekstra et al.)
- Distance measures between distributions (Diks et al.)
- Information theoretic dependence measures
Nonlinear dynamics and chaos

Dynamical system

- State space $\Omega$
- Time variable $t \in \mathbb{Z}$ (discrete) or $t \in \mathbb{R}$ (continuous)
- Evolution operator $\phi(x, t)$, defines a map

$$\phi^t : \Omega \to \Omega, \quad \phi^t(x) = \phi(x, t)$$

(flow over time $t$)

- (semi-)group properties

$$\phi^0 = \text{Id}, \quad \phi^s \circ \phi^t = \phi^{s+t}$$

$s, t \geq 0 \ (\forall s, t \in \mathbb{Z} \ \text{resp.} \ \mathbb{R} \ \text{for invertible dynamics})$
The logistic map

Example: logistic map

Map:

\[ x_t = f(x_{t-1}) = ax_{t-1}(1 - x_{t-1}), \quad 0 < a < 4 \]
Iterating the logistic map
Qualitative behaviour depends on $a$

(a) $a = 2$, (b) $a = 3.5$, (c) $a = 4$
Bifurcation diagram for logistic map
Sensitive dependence on initial conditions

Near states may either diverge or converge in the long-run.

Average growth rate of near trajectories characterized by the Lyapunov exponent.

Lyapunov exponent (for 1-dimensional maps)

\[
\lambda = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \log |f'(x_t)| , \quad \text{with } f'(x) = \frac{df(x)}{dx}
\]

If the system has a unique invariant measure, \( \mu \) say, \( \lambda \) is the same for \( \mu \)-almost all initial states \( x_0 \sim \mu \).
Lyapunov exponents

Lyapunov exponent - logistic map

The graph shows the Lyapunov exponents as a function of the parameter $a$. The exponents range from $-3.5$ to $0$ on the vertical axis, and $a$ varies from $2.6$ to $4$ on the horizontal axis. The graph illustrates the behavior of the logistic map over different values of $a$. The transition to chaos is evident as $a$ increases, with the exponents indicating the stability or instability of the system.
The Hénon map

Example: The Hénon map

\[ x_{t+1} = 1 - ax_t^2 + y_t \]
\[ y_{t+1} = bx_t \]

Default values: \( a = 1.4, \ b = 0.3 \)

Jacobian

\[
J = \begin{pmatrix} -2ax_n & 1 \\ b & 0 \end{pmatrix} \quad \text{det} \ J = -b
\]

\( \Rightarrow \) volume contracting for \( |b| \leq 1 \)
Phase plot – Hénon attractor
Bifurcation plot – Hénon map

Bifurcation diagram - Henon map
Largest Lyapunov exponent – Hénon map
The Lorenz attractor

Lorenz system

\[
\begin{align*}
\dot{x} &= \sigma(y - x) \\
\dot{y} &= (r - z)x - y \\
\dot{z} &= xy - bz
\end{align*}
\]

Default parameter values

\[
\begin{align*}
b &= \frac{8}{3}, \\
\sigma &= 10, \quad \text{(Prandtl number)} \\
r &= 28 \quad \text{(Rayleigh number)}
\end{align*}
\]
Lorenz butterfly

\[ \text{Time series} \quad \text{Nonlinear dynamics and chaos} \quad \text{Attractor reconstruction from time series} \]
The Rössler dynamical system

Rössler system

\[
\begin{align*}
\dot{x} &= -y - z \\
\dot{y} &= x + ay \\
\dot{z} &= b + z(x - c)
\end{align*}
\]

Standard parameter values

\[
\begin{align*}
a &= 0.15, \\
b &= 0.2, \\
c &= 10
\end{align*}
\]
Rössler attractor
Takens reconstruction theorem

The state space can be reconstructed from an observed scalar time series.

Main idea: starting from a given state, a unique time series pattern arises

Use consecutive data \((X_t, X_{t+\tau}, \ldots, X_{t+(m-1)\tau})\) to fix the state at time \(t\)

Under which conditions does this work? If it does, how should \(m\) and \(\tau\) be chosen?
Takens reconstruction theorem

Assumptions (continuous time case):

- Unknown dynamical system

\[
\frac{dy}{dt} = F(y_t), \quad y \in \Omega
\]

- Dynamics confined to finite dimensional compact subspace \( M \subset \Omega \)

- Observations are generated as \( x_t = h(y_t) \), at regularly spaced times \( t = t_0 + k\tau \), \( (k \text{ integer}) \) where \( h \) is a continuous measurement function

\[
h : M \to \mathbb{R}, \quad y \mapsto h(y)
\]

- For flows: some requirements on the time interval \( \tau \) (e.g. no periodic orbits with period \( k\tau \))
Takens reconstruction theorem

Continuous time: flow over time interval $\tau$ is described by map

$$\phi^\tau : \Omega \rightarrow \Omega, \quad \phi^t(y) = \phi(y, t)$$

From now, use same notation for discrete and continuous time, with map $F : \Omega \rightarrow \Omega$ describing the flow over a time lag $\tau$

Note:

$$\Phi_2 : y \mapsto (h(y), h(F(y)))$$

is a map from $\mathcal{M} \subset \Omega$ to the plane, $\mathbb{R}^2$
Reconstruction of the Hénon attractor
Takens reconstruction theorem

Generalisation:

$$\Phi_m : y \mapsto (h(y), h(F(y)), \ldots, h(F^{m-1}(y)))$$

is a map from $M \subset \Omega$ to $\mathbb{R}^m$

Takens reconstruction theorem

For smooth measurement functions $h$, and for $m$ sufficiently large, $\Phi_m$ generically is a smooth invertible map from $M$ to $\mathbb{R}^m$ with a smooth inverse (diffeomorphism)
Reconstructed attractor

Corollary of the reconstruction theorem

Delay vectors \((x_t, x_{t+\tau}, \ldots, x_{t+(m-1)\tau}) \Omega \in \mathbb{R}^m\) lie on a faithful image of the attractor (for \(m\) large enough)

In case there is a finite dimensional attractor, the delay vectors also lie on a finite dimensional set, the reconstructed attractor

Takens: \(m \geq 2 \dim M + 1\) suffices

This bound can be improved further (Sauer et al., 1991): \(m > 2 \text{boxdim } A\) (box-counting dimension of the attractor)

For estimation of \(\dim A\), only \(m \geq \dim A\) is required
Pespective

- Control theory: Takens’ theorem is an observability result (Aeyels)
- Generalisation of Whitney’s embedding theorem (to a restricted set of maps from $M$ to $\mathbb{R}^m$)
Part II

Correlation integrals
4 Correlation integrals
   • Correlation integrals

5 Dynamic invariants
   • GP algorithm

6 Correlation integrals for noisy data
   • Observational vs dynamic noise
   • Gaussian kernel correlation integrals
Characterising the reconstructed attractor

- Delay vectors \((X_t, X_{t+\tau}, \ldots, X_{t+(m-1)\tau})\) for stationary time series have a well-defined long-run distribution.
- Reconstruction has an associated measure \(\mu_{m,\tau}\), called the reconstruction measure.
- The dynamics has some properties that don’t depend on the representation of the dynamics (dynamic invariants).
- For instance, correlation dimension and correlation entropy are independent on the representation.
**Correlation integrals**

**Definition of the correlation integral**

The correlation integral of the reconstruction measure for embedding dimension $m$ is

$$C_m(r) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} \Theta(r - \|x - y\|) \mu(dy) \mu(dx)$$

with

$$\Theta(s) = \begin{cases} 
0, & s < 0, \\
1, & s \geq 0.
\end{cases}$$

(Heaviside function)

**Compact notation**

$$C_m(r) = \int_{\mathbb{R}^m} \int_{B_r(x)} \mu(dy) \mu(dx) := \int_{\mathbb{R}^m} \mu(B_r(x)) \mu(dx)$$
Scaling law

For finite dimensional attractors: scaling relation

\[ C_m(r) \sim e^{-mrK_2} r^{D_2} \]

- \( D_2 \) is called the correlation dimension. Geometric measure of complexity
- \( K_2 \) is known as the correlation entropy. Dynamical measure of complexity
Renyi spectrum

Renyi family of correlation integrals indexed by order $q$

$$C_{q,m}(r) = \left( \int_{\mathbb{R}^m} \left( \int_{\mathbb{R}^m} \Theta(r - \|x - y\|) \mu(dy) \right)^{q-1} \mu(dx) \right)^{\frac{1}{q-1}}$$

$$= \left( \int_{\mathbb{R}^m} \left( \mu(B_r(x)) \right)^{q-1} \mu(dx) \right)^{\frac{1}{q-1}}$$

Scaling law

$$C_{q,m}(r) \propto e^{-mrK_q} r^{D_q}$$

Directly generalises $D_2$ and $K_2$
Sufficient condition for chaos

Definition of chaos: $K_1 > 0$

Estimation of inner integral

$$\left( \int_{\mathbb{R}^m} \Theta(r - \|x - y\|) \mu(dy) \right)^{q-1}$$

problematic for $q = 1$. For $q \to 1$, one finds (l’Hopital)

$$C_{1,m}(r) := \lim_{q \to 1} C_{q,m}(r) = \int \ln \left( \int_{B_r(x)} \mu(dy) \right) \mu(dx)$$

$K_2 > 0$ implies $K_1$, hence chaos
Estimating correlation integrals

**empirical reconstruction measure**

The set of delay vectors with equal mass $\frac{1}{n}$ associated to each point (also called empirical delay vector distribution)

**Empirical correlation integral**

\[
\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Theta(r - \|x_i - x_j\|)
\]

(V-statistic)

\[
\frac{2}{n(n-1)} \sum_{i=2}^{n} \sum_{j=1}^{i} \Theta(r - \|x_i - x_j\|)
\]

(U-statistic)
Theiler correction

In practice we use

\[
\frac{2}{(n - T)(n - T + 1)} \sum_{i=T+1}^{n} \sum_{j=1}^{i-1} \Theta(r - \|x_i - x_j\|)
\]

\(T\) is called the Theiler correction. \(T = 1\) corresponds to a U-statistic (no Theiler correction)

Statistically, the Theiler correction is a finite sample size correction
Grassberger-Procaccia method for estimating $D_2$ and $K_2$

1. Estimate the correlation integral for $m = 1, \ldots, m_{\text{max}}$

$$\hat{C}_m(r) = \frac{2}{(n-T)(n-T+1)} \sum_{i=T+1}^{n} \sum_{j=1}^{i-1} \Theta(r - \| x_i - x_j \|)$$

for a range of $r$-values, typically $r = ca^k$ for integer $k$

2. Look for a ‘scaling region’ of $r$-values

3. Estimate $D_2$ and $K_2$ using the estimated correlation integrals from the scaling region

4. Check for convergence with $m$
Correlation integrals for the Rössler
Estimated correlation dimension and entropy
Application to atrial fibrillation
Atrial fibrillation: correlation integrals

Correlation integrals
Dynamic invariants
Correlation integrals for noisy data

Graph showing correlation integrals for different values of m (2 and 16). The x-axis represents \( \log_2 r \) and the y-axis represents \( \log_2 C(r) \). The slope of the curves varies with different values of m.
Atrial fibrillation: estimated $D_2$ and $K_2$
Observational vs dynamic noise

**Observational noise**

\[ Y_t = F(Y_{t-1}) \]
\[ X_t = h(Y_t) + \varepsilon_t, \]

**Dynamic noise**

\[ Y_t = F(Y_{t-1}, \varepsilon_t), \]

For example:

\[ Y_t = F(Y_{t-1}) + \varepsilon_t, \]
\[ X_t = h(Y_t) \]
Phase plots from clean and noisy (5%, normal) Rössler data
Correlation integrals for noisy Rössler data

- Graph showing the correlation integrals for noisy data with different values of $m$.
- The upper graph displays the slope ($\log_2(r)$) on the y-axis and the correlation integral on the x-axis.
- The lower graph shows the log-log plot of $\log_2 C(r)$ against $\log_2(r)$.
- The graph highlights the behavior of the correlation integrals for $m=2$ and $m=16$.
Effect of observational noise on reconstruction measure

Delay vector \((m\text{-dimensional})\)

\[
X^m_t := (X_{t-m+1}, \ldots, X_t)
\]

is replaced by

\[
X^m_t := (X_{t-m+1} + \varepsilon_{t-m+1}, \ldots, X_t + \varepsilon_t)
\]

⇒ The ‘clean’ reconstruction measure \(\mu^0_m\) is replaced by

\[
\mu_m = \mu^0_m \circ \nu_m \text{ (convolution)}
\]

where \(\nu_m\) denotes the noise reconstruction measure
Correlation integrals under observational noise

Correlation integral

\[ C_m(r) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} l_{[0,r]}(\|x - y\|) \mu(dx) \mu(dy) \]

\[ = \int_{\mathbb{R}^m} \cdots \int_{\mathbb{R}^m} l_{[0,r]}(\|x + v - y - w\|) \nu(dv) \nu(dw) \mu^0(dx) \mu^0(dy) \]

\[ = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} l_{[0,r]}(\|z + s\|) \nu(ds) \xi(dz) \]

with

\[ \eta(A) = \mathbb{P}[V - W \in A], \quad (V, W \sim \nu \text{ independent}) \]

and

\[ \xi(A) = \mathbb{P}[X - Y \in A], \quad (X, Y \sim \mu^0 \text{ independent}) \]
Correlation integrals under observational noise (ctd)

Clean correlation integral:

\[ C^0_m(r) = \int_{\mathbb{R}^m} I_{[0,r]}(\|z\|) \xi(dz) = \int_0^r dC^0_m(s) \]

Correlation integral in presence of observational noise

\[ C_m(r) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} I_{[0,r]}(\|z + s\|) \nu(ds) \xi(dz) \]

Gaussian observational noise already difficult analytically

- Inner integral difficult to evaluate
- Generally, noisy CI is not a functional of the ‘clean’ CI
Gaussian kernel correlation integral

Although the CI generally is not a functional of the ‘clean’ CI, it is if $\| \cdot \|$ is the Euclidean norm. In that case we can define:

$$\int_{\mathbb{R}^m} l_{[0,r]}(\| z + s \|) \nu_m(ds) := g_r(\| z \|),$$

and hence

$$C_m(r) = \int_{\mathbb{R}^m} g_r(\| z \|) \xi(dz) = \int_0^\infty g_r(s)f^0_{\| z \|}(s) ds = \int_0^\infty g_r(s)dC^0_m(s)$$

More generally, whenever the kernel function used for calculating the CI depends on the Euclidean distance, the noisy CI is a functional of the noise-free CI.
Relations between ‘noisy’ and ‘clean’ correlation integrals

The ‘classical’ correlation integral is of the general form

\[ T_m^0(h) = \int_0^\infty K_{m,h}(s)f_{\|Z\|}^0(s) \, ds \]

with \( K_{m,h}(s) = I_{[0,h]}(s) \).

\[ T_m(r) = \int_0^\infty g_r(s)f_{\|Z\|}^0(s) \, ds \]

If \( g_r(s) \) is of the same form as the kernel function used for calculating the correlation integral, the noisy and clean CI’s then will be automatically simply related.
The Gaussian kernel

For

$$K_{m,h}(s) = e^{-\frac{1}{2} \frac{s^2}{h^2}}$$

$$g_r(||z||) = \int_{\mathbb{R}^m} e^{-\frac{1}{2h^2} ||z+s||^2} \nu_m(ds)$$

$$= (2\pi \sigma)^{\frac{m}{2}} \int_{\mathbb{R}^m} e^{-\frac{1}{2h^2} ||z+s||^2} - \frac{1}{2\sigma^2} (s_1^2 + \cdots + s_m^2) \, ds_1 \cdots ds_n$$

$$= \prod_{i=1}^{m} \int_{\mathbb{R}} e^{-\frac{1}{2h^2} (z_i+s_i)^2} - \frac{1}{2\sigma^2} s_i^2 \, ds_1 \cdots ds_n$$

Since

$$\int_{\mathbb{R}} e^{-\frac{1}{2h^2} (z+s)^2} - \frac{1}{2\sigma^2} s^2 = \left( \frac{h^2}{h^2 + \sigma^2} \right)^{\frac{1}{2}} e^{-\frac{1}{2(h^2+\sigma^2)}} \, ds$$

one finds

$$g_r(||z||) = \left( \frac{h^2}{h^2 + \sigma^2} \right)^{\frac{m}{2}} e^{-\frac{||z||^2}{2(h^2+\sigma^2)}}$$
Modified scaling law in the presence of noise

Gaussian kernel correlation integral in the presence of noise

\[ T_m(h) = \left( \frac{h^2}{h^2 + \sigma^2} \right)^{\frac{m}{2}} T_m^0(h^2 + \sigma^2) \]

The noise-free gaussian kernel behaves similar to the usual

\[ T_m^0(h) \simeq \text{cnst} \times m^{-\frac{D}{2}} e^{-Km h^D}. \]

Modified scaling law

\[ T_m(h) \simeq \text{cnst} \times m^{-\frac{D}{2}} \times \left( \frac{h^2}{h^2 + \sigma^2} \right)^{\frac{m}{2}} e^{-Km(h^2 + \sigma^2)^{\frac{D}{2}}} \]
Noisy and noise-free Hénon CI

\[ \log_2 \tilde{T}_m(h) \]

\( m = 1 \)

\( m = 10 \)
Estimation of $D_2$ and $K_2$ in presence of noise

Nonlinear least squares (Levenberg-Marquardt), either

- Weighted least squares. Requires standard errors and possibly covariances of $\hat{T}_m(h)$.
- We used $\text{Var} \hat{T}_m(h) \propto \hat{T}_m(h)(2 - \hat{T}_m(h))$.

- Unweighted
- Implicitly also imposes weights (log-log scale, lin-lin scale)
Estimated invariants from noisy Hénon CI
Course-grained quantities

Even if assumptions made so far (determinism + gaussian observational noise) are violated, one may use correlation integrals.

Idea is to use behavior of correlation integrals as a measure of complexity:

- Large slope of $\ln C_m(r)$ indicates large number of relevant state variables
- Difference $\ln C_m(r) - \ln C_{m+1}(r)$ is a measure of unpredictability at scale $r$

Motivates examining course-grained quantities
Definition of course-grained quantities

### Course-grained correlation dimension and entropy

**Coarse-grained correlation dimension**

\[
D_2(m, r) = \frac{d \ln C_m(r)}{d \ln r}
\]

**Coarse-grained correlation entropy**

\[
K_2(m, r) = \ln C_m(r) - \ln C_{m+1}(r)
\]
Part III

Local linear prediction
Prediction by analogy

Casdagli method
Prediction by analogy, nearest neighbour method

Reconstruct state space, state vectors $X^m_s$, $1 \leq s \leq n$

Idea is to exploit

$$X_{t+\ell} = g(X^m_t)$$

To make $\ell$-step-ahead forecast from $X^m_t$ we need an approximation (estimate) of $g$

- find $k$ nearest neighbours $X^m_{s_i}$ among $X^m_s$, $s < t$ or $s \neq t$
- locally constant model leads to prediction by averaging

$$\hat{X}_{t+\ell} = \sum_{i=1}^{k} X_{s_i+\ell}$$

- alternatively, construct local linear map through linear regression (requires $k > m$). Model:

$$X_{t+\ell} = a_0 + \sum_{j=1}^{m} a_j X_{s_i+1-j} + \varepsilon_{t+\ell}, \quad i = 1, \ldots, k$$
Prediction by analogy, kernel methods

Comparable to $k$-nearest neighbour method, only using all points with weights determined by distance

Prediction

$$\hat{X}_{t+\ell} = \sum_s w_{s,t} X_{s+\ell}$$

weights $w_i$ determined by distance in state space

$$w_{s,t} = \frac{K_h(X^m_s - X^m_t)}{\sum_s K_h(X^m_s - X^m_t)}.$$
Prediction by analogy, other methods

- Polynomials (global nonlinear model)
- Neural networks (global nonlinear)
- Radial basis functions (local, linear in coefficients)

\[ F(x) = \alpha_0 + \sum_i \alpha_i \Phi(||x - x_i||) \]

e.g. \( \Phi(s) = 1/(1 + e^{bx-c}) \)

Parameter estimation by error backpropagation (gradient descent)
Casdagli method

- Divide the data in a fitting set \( x_1, \ldots, x_{N_f} \) and a testing set \( x_{N_f+1}, \ldots, x_{N_f+N_t} \)
- Vary \( k \) in the nearest neighbour method, for each \( k \):
  - Choose a number of random reference points to predict from
  - Determine prediction error \( e_i(k) = |\hat{x}_{i+\ell}(k) - x_{i+\ell}| \)
  - Repeat to determine RMSE

\[
E_m(k) = \left( \frac{1}{n} \sum_i e_i^2(k) \right)^{1/2} / \sigma
\]
Example Casdagli method (Hénon data)

\[ m = 3, \varepsilon = 0.1, \text{200 predictions} \]
Prediciton by analogy

- Takens’ reconstruction theorem, correlation integrals, fractal dimensions, correlation entropy, estimation of dynamic invariants, local linear prediction
- Nonparametric tests based on correlation integrals: Divergences between reconstruction measures, U-statistics estimators, tests for symmetry of multivariate distributions, attractor comparison, testing for reversibility, tests for serial independence and linearity
- Statistical aspects of nonparametric tests: asymptotic results for U-statistics in time series context, bootstrap and Monte Carlo tests, consistency, bandwidth selection problem, diagnostic model checking, nuisance parameters, tests based on empirical copulas
- Granger causality tests: conditional independence, linear versus nonlinear Granger causality, nonparametric Granger causality tests, testing for Granger causality using correlation integrals, consistency, local measures of conditional dependence