Time-Series Analysis with Linear Methods

Gerrit Ansmann

Motivation

"I wanna do non-linear time-series analysis
– why care about linear methods?"

- Linear methods yield complementary, useful information.
- Non-linear methods may be overkill.
- Linear methods may decide about prerequisites for non-linear methods.
- Some linear methods are basic ingredients of non-linear methods.
- Get acquainted with the pitfalls of data analysis.

The Distribution of Values

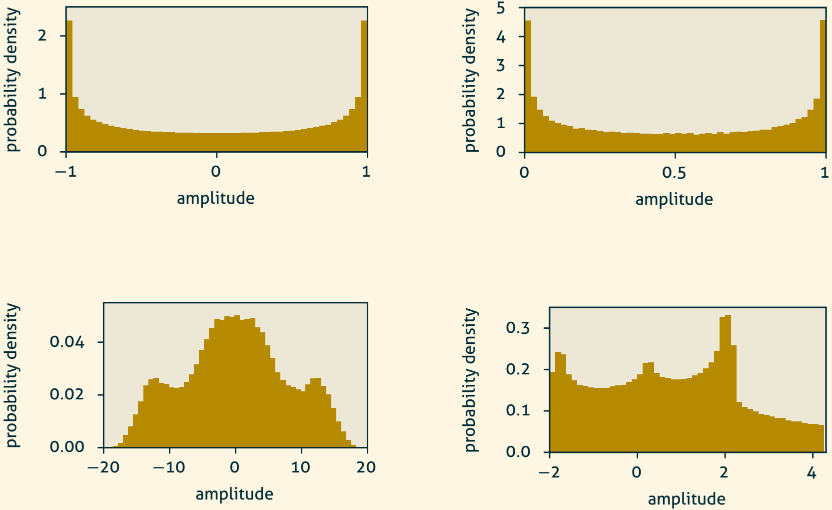
Assumption

Each value of the time series is independently sampled from some distribution.

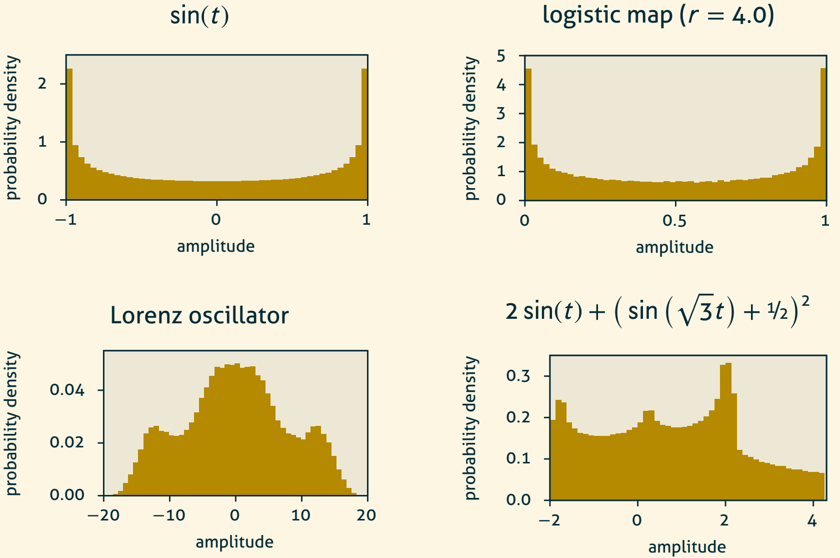
Assumption implies:

- No memory.
- No dynamics.
- Time is not important.
- Stationarity (later).

Distribution of Values: Examples



Distribution of Values: Examples



Statistical Moments 1: The Mean

Setup: time series x: $x_1, x_2, \ldots, x_N$

First moment: mean

$$\bar{x} := \frac{1}{N} \sum_{t=1}^{N} x_t$$

Statistical Moments 1: The Mean

Setup: time series x: $x_1, x_2, \ldots, x_N$

First moment: mean

$$\bar{x} := \frac{1}{N} \sum_{t=1}^{N} x_t$$

mean vs. expected value:

- The mean ($\bar{x}$) is a property of a dataset.
- The expected value ($\langle x \rangle$) is a property of a population.
- If a dataset is sampled from some poplation,
  $\bar{x}$ is the best estimator for $\langle x \rangle$ (of that population).
  (Law of large numbers)

Statistical Moments 2: The Variance

Second moment: variance

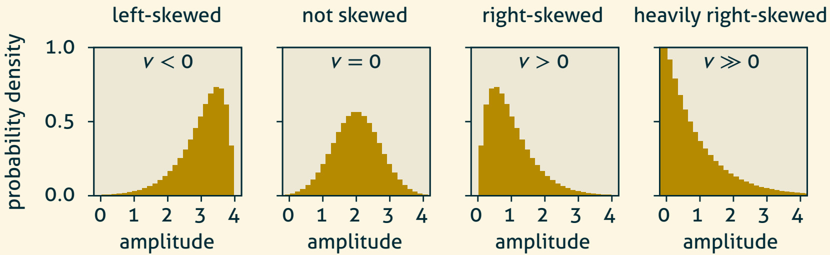
$$\sigma(x)^2 := \frac{1}{N-1} \sum_{t=1}^{N} \left(x_t - \bar{x}\right)^2$$

- width of the distribution, variability of the time series
- σ: standard deviation
- normalisation factor:
  - $N-1$: estimating variance from a dataset
  - N: variance of a population

Statistical Moments 3: The Skewness

Third moment: skewness

$$v(x) := \frac{1}{N} \sum_{t=1}^{N} \left(\frac{x_t - \bar{x}}{\sigma(x)} \right)^3$$

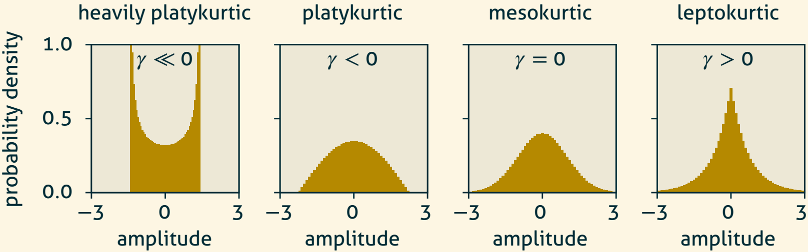


- $v(x) = 0$ for any symmetric distribution

Statistical Moments 4: The Kurtosis

Fourth moment: (excess) kurtosis

$$\gamma(x) := \frac{1}{N} \sum_{t=1}^{N} \left(\frac{x_t - \bar{x}}{\sigma(x)} \right)^4 - 3$$



- the normal distribution has $\gamma = 0$

Interpreting Skewness and Kurtosis

- Typical noise is a superposition of many small effects.

→ Typical noise is approximatively normally distributed (central limit theorem)

- The normal distribution is symmetric and mesokurtic.

- A significantly non-zero skewness and kurtosis hint at "something going on", e.g.:
  - non-linearity of measurement
  - dynamics
  - non-linear dynamics

Statistical Tests

Example: skewness test

- *assumption / prerequisite:*
  data independently sampled from some population.
- *null hypothesis:*
  population not skewed
- *p value / error probability / significance:*
  probability to find observed skewness
  in a population complying with the null hypothesis
  $\rightsquigarrow$ probability that null hypothesis is true

Typical procedure:

1. Choose significance threshold α, e.g., $\alpha = 0.05$.
2. If $p < \alpha$, reject null hypothesis, e.g., consider data skewed.

Beware the Prerequisites

Significance values are meaningless
if assumptions are not fulfilled.

Example: Results of skewness test for $\{\sin(t) \mid t \in \mathscr{T}\}$:

$\mathscr{T}$	p
$(0.00, 0.01, \ldots, 9.00)$	$4 \cdot 10^{-9}$
$(0.00, 0.01, \ldots, 40.00)$	0.02
$(0.00, 0.01, \ldots, 41.00)$	0.002
$(0.0, 0.1, \ldots, 9.0)$	0.05
$(0, 1, \ldots, 100)$	0.95

Problem: Data not idependent.

The Kolmogorov–Smirnov (KS) Test

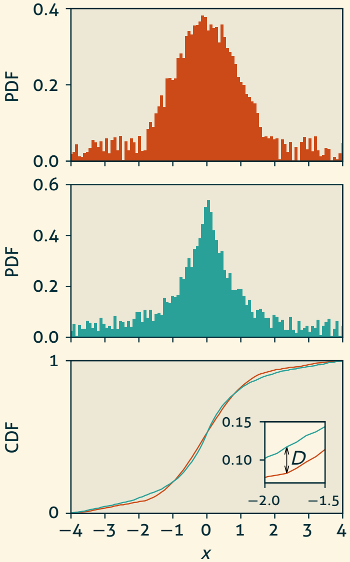
- Based on cumulative distribution functions:

$$\mathrm{CDF}(x) := \int_{-\infty}^{x} \mathrm{PDF}(\tilde{x})\,\mathrm{d}\tilde{x}$$

- Obtain p from maximal difference between CDFs:

$$D := \max_x \left|\mathrm{CDF}_1(x) - \mathrm{CDF}_2(x)\right|$$

- D independent from several scales



Beware the Prerequisites (Once More)

Results of Kolmogorov–Smirnov test for comparing $\{\sin(t) \mid t \in \mathcal{T}_1\}$ with $\{\sin(t) \mid t \in \mathcal{T}_2\}$:

$\mathcal{T}_1$	$\mathcal{T}_2$	p
$(0.00, 0.01, \ldots, 9.00)$	$(3.00, 3.01, \ldots, 12.00)$	$6 \cdot 10^{-33}$
$(0.00, 0.01, \ldots, 40.00)$	$(3.00, 3.01, \ldots, 43.00)$	$2 \cdot 10^{-5}$
$(0.0, 0.1, \ldots, 9.0)$	$(3.0, 3.1, \ldots, 12.0)$	0.001
$(0, 1, \ldots, 100)$	$(3, 4, \ldots, 103)$	0.99

Problem: Data not idependent.

Pearson's Correlation Coefficient

Setup: two time series $x_1, x_2, \ldots, x_N$ and $y_1, y_2, \ldots, y_N$.

Covariance

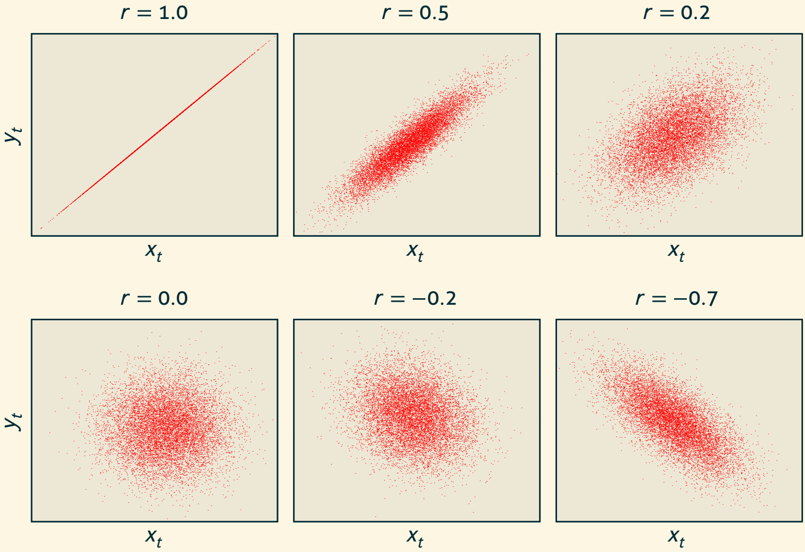
$$\operatorname{cov}(x,y) := \frac{1}{N-1} \sum_{t=1}^{N} \left(x_t - \bar{x}\right)\left(y_t - \bar{y}\right)$$

Pearson's *r*

$$r(x,y) := \frac{\operatorname{cov}(x,y)}{\sigma(x)\,\sigma(y)}$$

- $r = 1$: perfect correlation
- $r = 0$: no correlation
- $r = -1$: perfect anti-correlation

Pearson's Correlation Coefficient



Cross-Correlation

Motivation:

- Possible offset in time data.
- Sensors may capture dynamics with delay between them.

Shifted time series

$$x^{\tau} \ : \ x_{t+\tau}, x_{t+\tau+1}, x_N$$

Cross-correlation

$$C\,(x,y,\tau) = r\left(x,y^{\tau}\right)$$

(with appropriately truncated time series)

- Symmetry: $C(x,y,\tau) = C(y,x,-\tau)$

Application of Cross-Correlation

Find delay and synchrony between two time series x and y:

1. Find $\hat{\tau}$ that maximises cross-correlation:

$$\hat{\tau} := \operatorname{argmax}_\tau C(x,y,\tau)$$

2. Use $C(x,y,\hat{\tau})$ as a measure for synchrony.

Problems:

- Assumes comparable dynamics.
- Assumes simple synchronisation.

The Autocorrelation

Autocorrelation

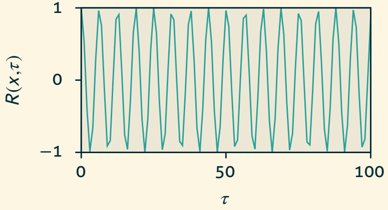
$$R(x,\tau) := C(x,x,\tau) = r(x,x^{\tau})$$

(with appropriately truncated time-series)

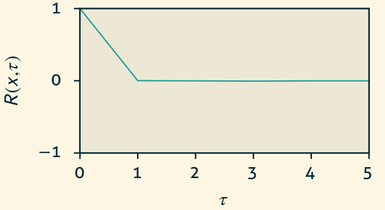
- $R(x,\tau) = R(x,-\tau)$
- $R(x,0) = 1$
- Positive autocorrelation implies some repetition.

Autocorrelation: Examples

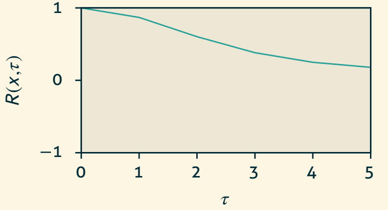
$\sin(t)$



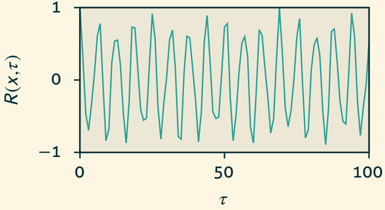
logistic map ($r = 4.0$)



Lorenz oscillator



$2\sin(t) + \left(\sin\left(\sqrt{3}t\right) + 1/2\right)^2$



Rank-Based Methods

Consider how values rank instead of actual values.

- Advantage: robust against outliers
- Advantage: often fewer constraints on data
- Disadvantages: information is discarded

amplitude-based method	rank-based analoga
mean	median
Pearson's r	Kendall's τ, Spearman's ρ
Kolmogorov–Smirnov test	Mann–Whitney test

Stationarity

Stationarity

A process is called *stationary* if the distribution of its states over an ensemble of realisations of that process does not depend on time.

This implies constant statistical moments:

- mean
- variance
- skewness
- kurtosis
- …

and mixed moments:

- covariance
- …

Stationarity

Stationarity

A process is called *stationary* if the distribution of its states over an ensemble of realisations of that process does not depend on time.

Examples for non-stationary processes:

- dynamics with changing parameters
- driven dynamics
- transient dynamics

Stationarity

Stationarity

A process is called *stationary* if the distribution of its states over an ensemble of realisations of that process does not depend on time.

- prerequisite of most analysis techniques
- ensures reproducability
- required for ergodicity
- depends on the time scale:
  - On short time scales, an non-stationary process can be approximated as stationary.
  - On long time scales, instationarities may be regarded as parts of the dynamics or a driver.

Stationarity

Stationarity

A process is called *stationary* if the distribution of its states over an ensemble of realisations of that process does not depend on time.

Weak stationarity

A process is called *weakly stationary* if its mean, variance, and covariances do not depend on time.

Frequency Spectrum

Assumption

The time series can be decomposed into periodic components.

Assumption implies:

- Periodicity, quasiperiodicity.
- No chaos.
- Memory.

The Fourier Transform

Continuous Fourier transform

$$\hat{x}(\nu) := \int_{-\infty}^{\infty} x(t) \exp(-2\pi i \nu t) \, \mathrm{d}t$$

Discrete Fourier transform

$$\hat{x}(k) := \sum_{t=0}^{N} x_t \exp\left(\frac{-2\pi i k t}{N}\right)$$

Numerical realisation:

- Fast Fourier Transform (FFT)
- Beware how the output is aligned.

Properties of the Fourier Transform

Convolution theorem: $\widehat{x * y} = \hat{x} \cdot \hat{y}$

Correlation theorem: $\widehat{C(x,y)} = \hat{x}^* \cdot \hat{y}$

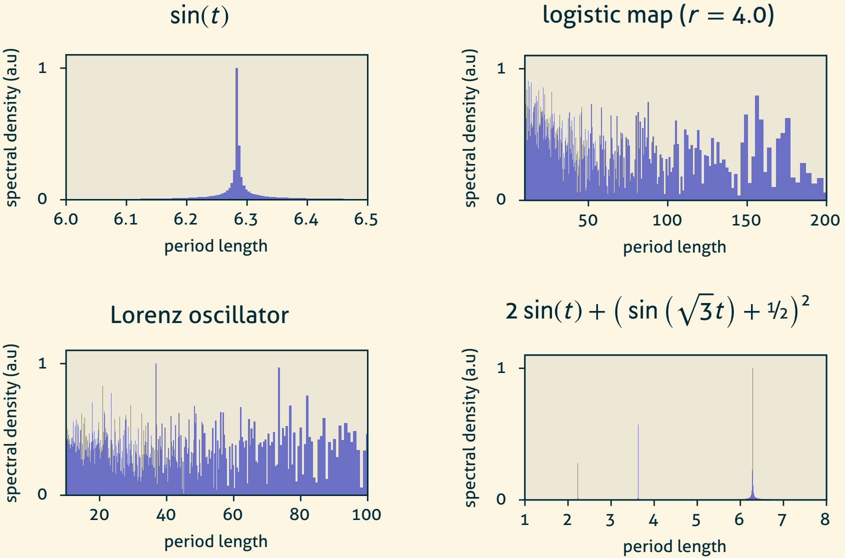
Wiener-Khinchin theorem: $\widehat{R(x)} = \widehat{C(x,x)} = \hat{x}^* \cdot \hat{x} = |x|^2$

Plancherel theorem: $\sum_{t=1}^{N} x_t^* \cdot y_t \propto \sum_{k=1}^{N} \hat{x}_k^* \cdot \hat{y}_k$

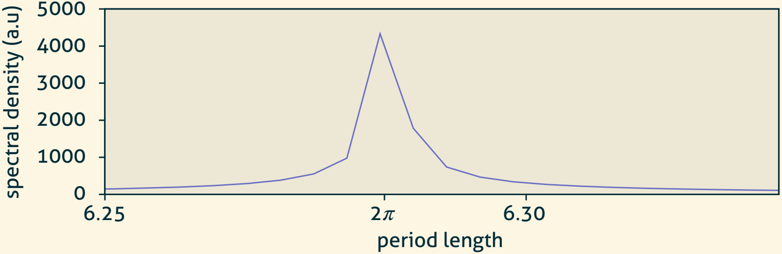
Parseval's theorem: $\sum_{t=1}^{N} \left|x_t\right|^2 \propto \sum_{k=1}^{N} \left|\hat{x}_k\right|^2$

... and respective analogues for the inverse Fourier transform

Fourier Transform: Examples



Spectral Leakage

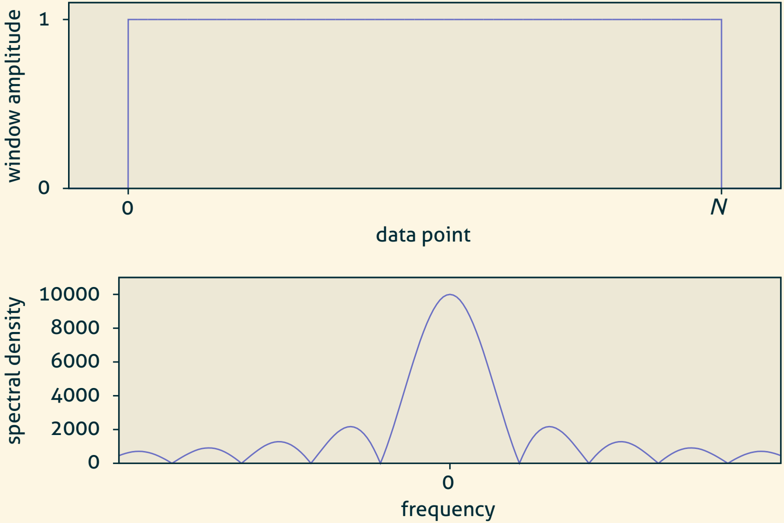


Problem: real signal is inevitably truncated: $x_t = p_t \cdot \chi_{[1,N]}(t)$

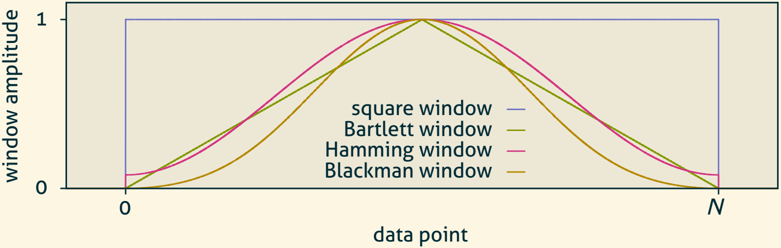
- p_t: "eternal" periodic signal.
- $\chi_{[1,N]}(t) := \begin{cases} 1 & \text{if } 1 \leq t \leq N \\ 0 & \text{otherwise} \end{cases}$: window

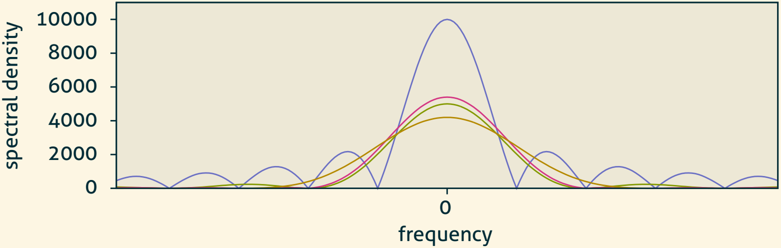
$\rightarrow \widehat{\chi_{[1,N]}}$ is convoluted onto $\hat{p}$ (convolution theorem).

Spectral Leakage – Windows

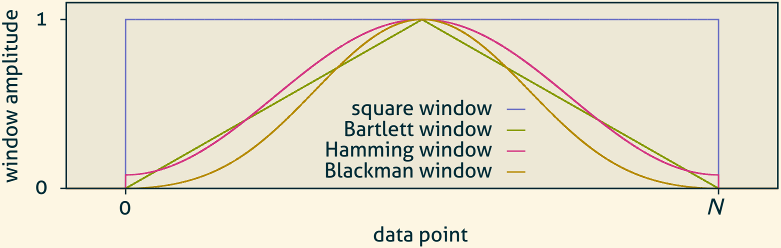


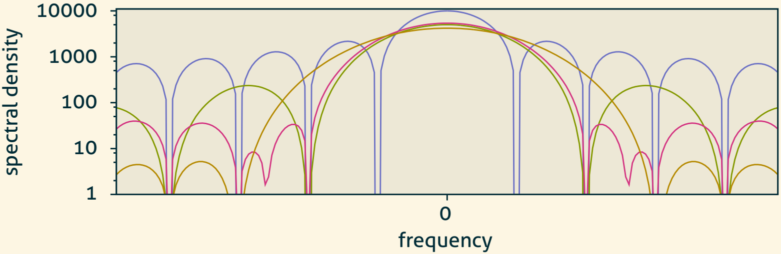
Spectral Leakage – Windows



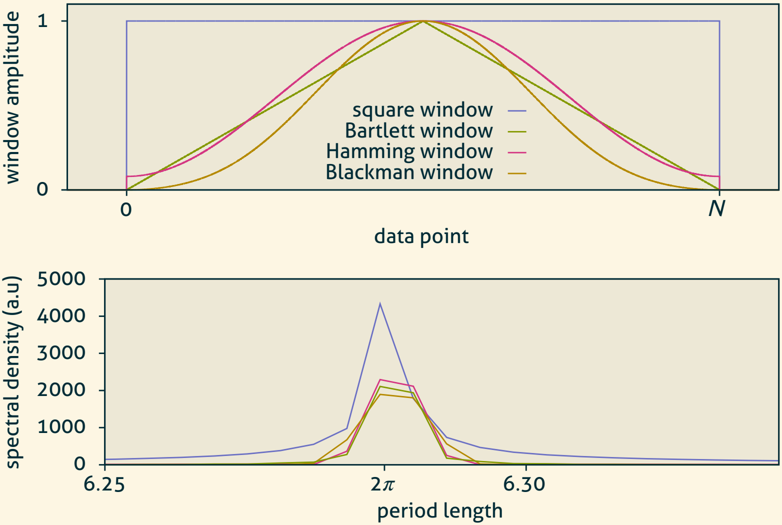


Spectral Leakage – Windows





Spectral Leakage – Windows



Uncertainty of the Fourier Transform

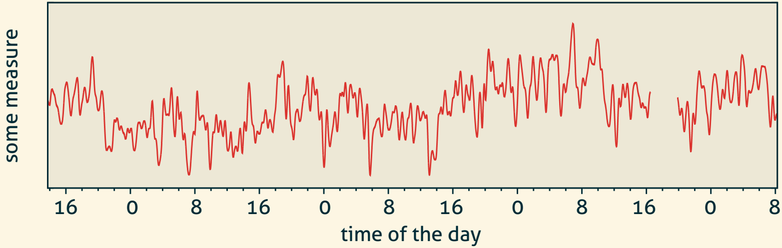
$$\sigma(\hat{x}(\nu)) \approx \langle \hat{x}(\nu) \rangle$$

Standard deviation is as large as the actual value.

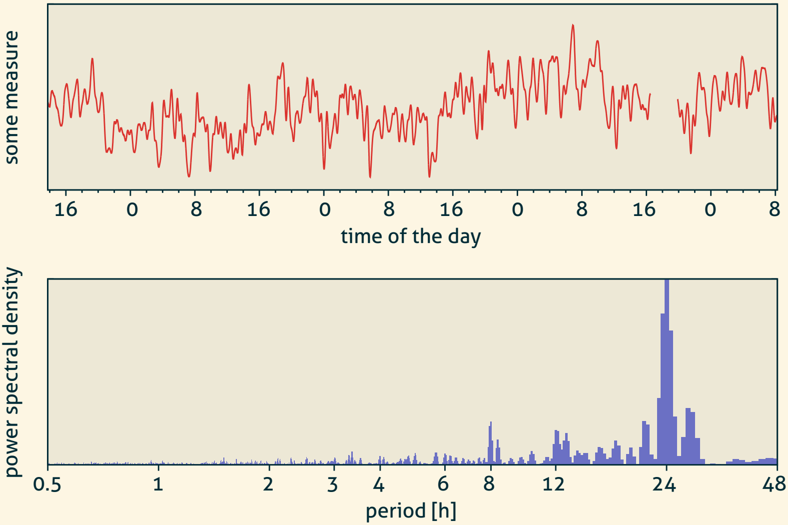
→ Averaging over moving windows in the time domain.

→ Moving average in the frequency domain.

A Real Example



A Real Example



Linear Stochastic Processes

Processes whose realisations depend on chance.

- demonstrate limits of linear methods
- contain most real linear processes as a special case
- null model / null hypothesis
- used for data-driven modelling and forecasting

White Noise

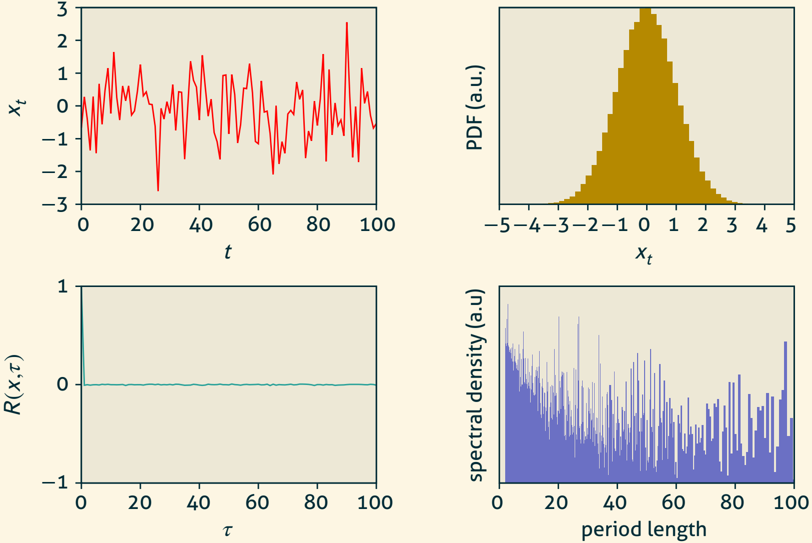
White noise

Each sample/value is independently sampled
from the same distribution:

$$x_t = \varepsilon_t$$

- All frequencies are equally present.
- Autocorrelation is zero, except for a delay of 1.
- Most often: white Gaußian noise.
- Basis for the following models.

Gaussian White Noise



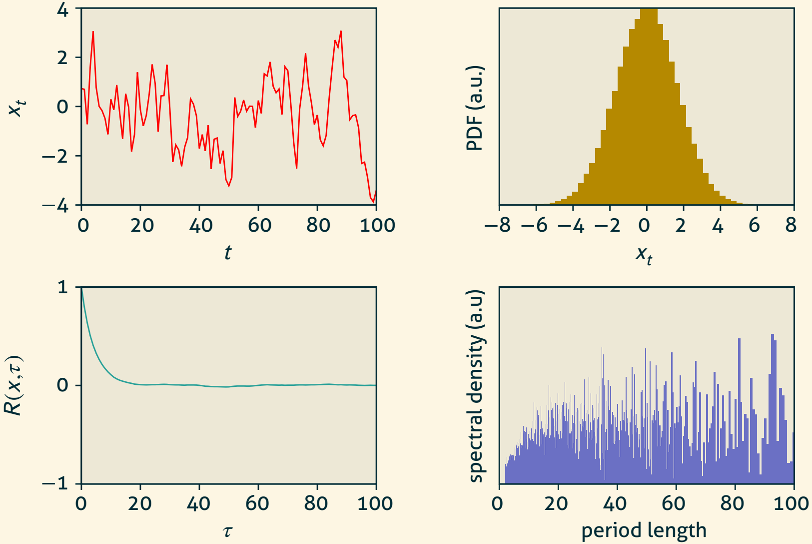
Autoregressive Processes of Order 1

Autoregressive process of order 1 (AR(1))

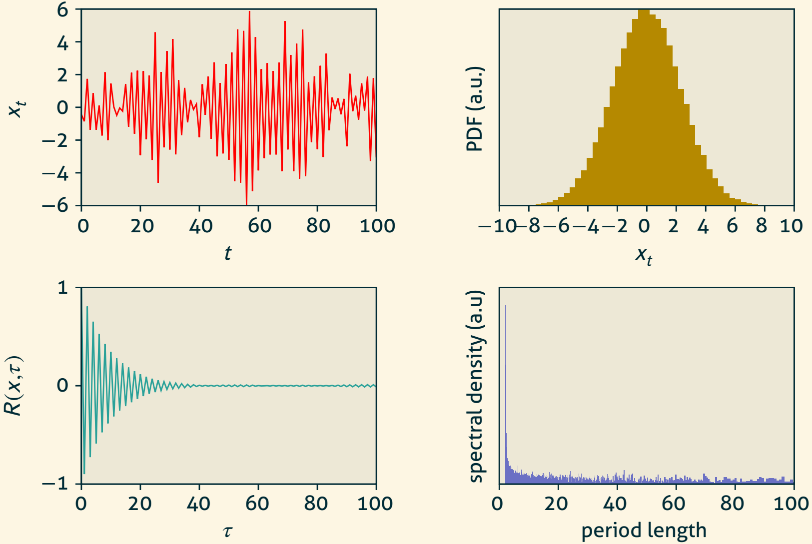
$$x_t = ax_{t-1} + \varepsilon_t$$

- Idea: Random process with some memory.
- Autocorrelation decays exponentially ($a > 0$) or exponentially damped oscillation ($a < 0$).

AR(1) Process ($a = 0.8$)



AR(1) Process ($a = -0.9$)



Autoregressive Processes

Autoregressive process of order p (AR(p))

$$x_t = \sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t$$

- Idea: Random process with some more memory.
- Autocorrelation is superposition of exponential decays and exponentially damped oscillations.

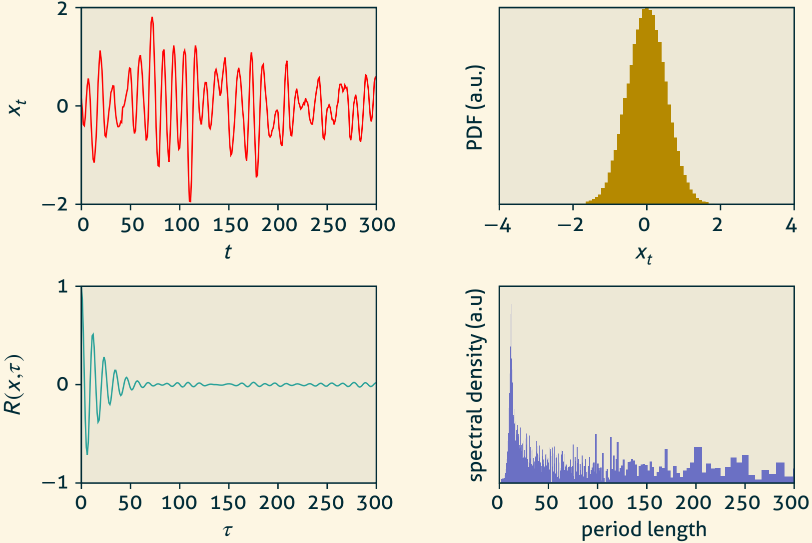
ARMA Processes

Autoregressive-moving-average process (ARMA(p,q))

$$x_t = \sum_{i=1}^{p} a_i x_{t-i} + \sum_{j=0}^{q} b_j \varepsilon_{t-j}$$

- Typically: $b_0 = 1$
- Idea: Random process with memory and smoothed noise.

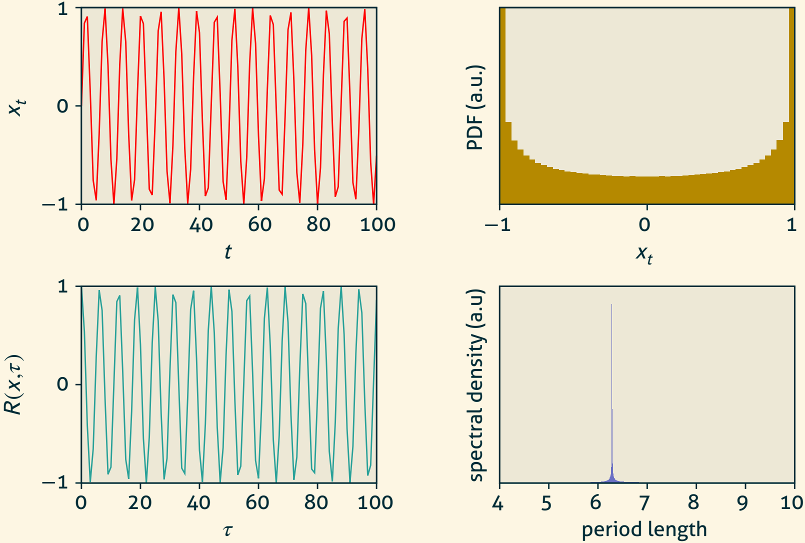
ARMA(4,4) Process



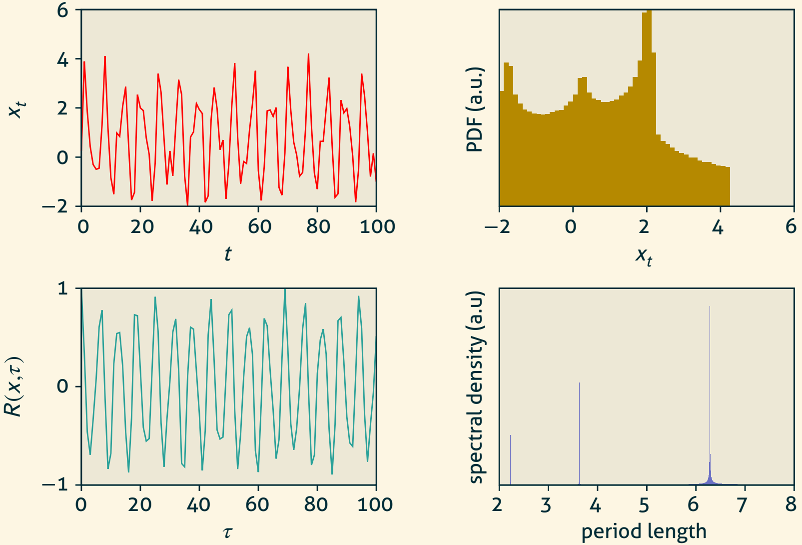
Further Stochastic Processes

- continuous-time, e.g., stochastic differential equations
- non-linear stochastic processes

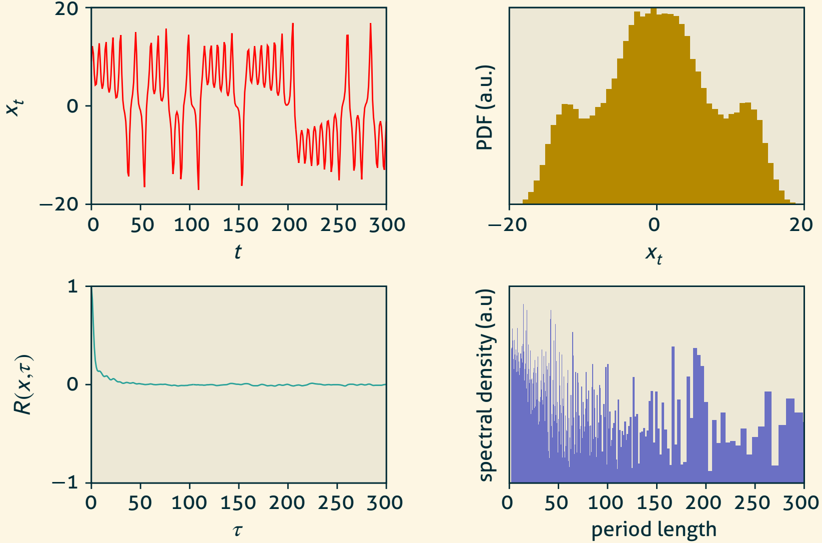
Sine Wave


1
0
−1
0
20
40
60
80
100
x_t
t
PDF (a.u.)
−1
0
1
x_t
1
0
−1
0
20
40
60
80
100
$R(x,\tau)$
τ
spectral density (a.u)
4
5
6
7
8
9
10
period length

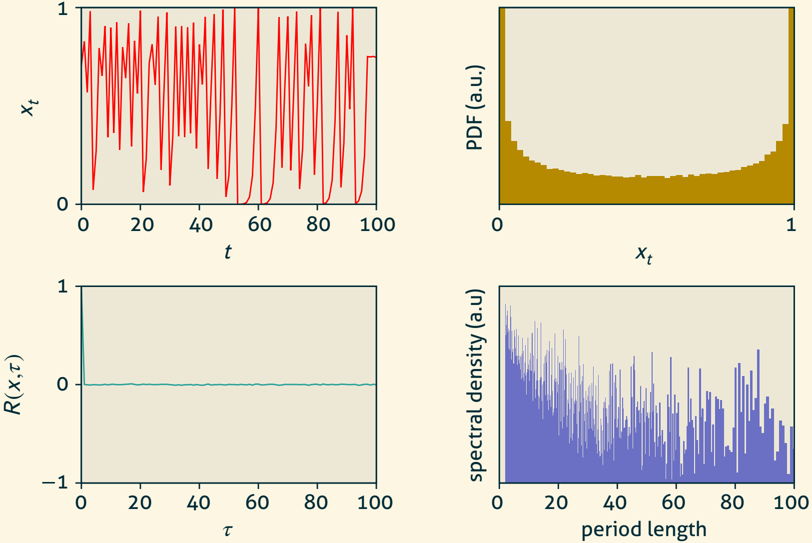
Quasiperiodic



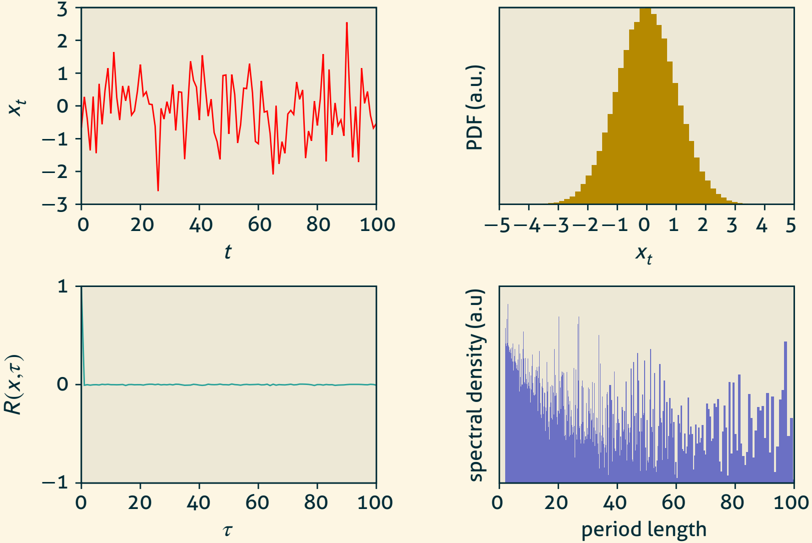
Lorenz Oscillator



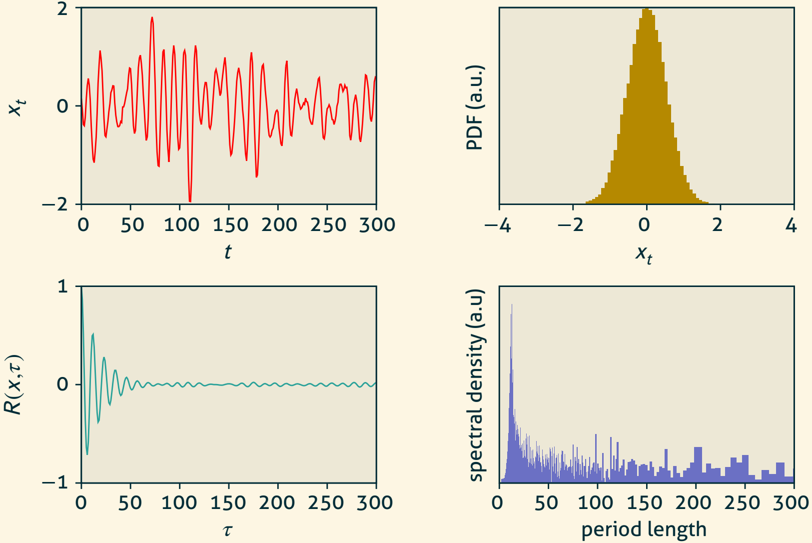
Logistic Map



Gaussian White Noise



ARMA(4,4) Process



Capabilities and Restrictions of Linear Methods

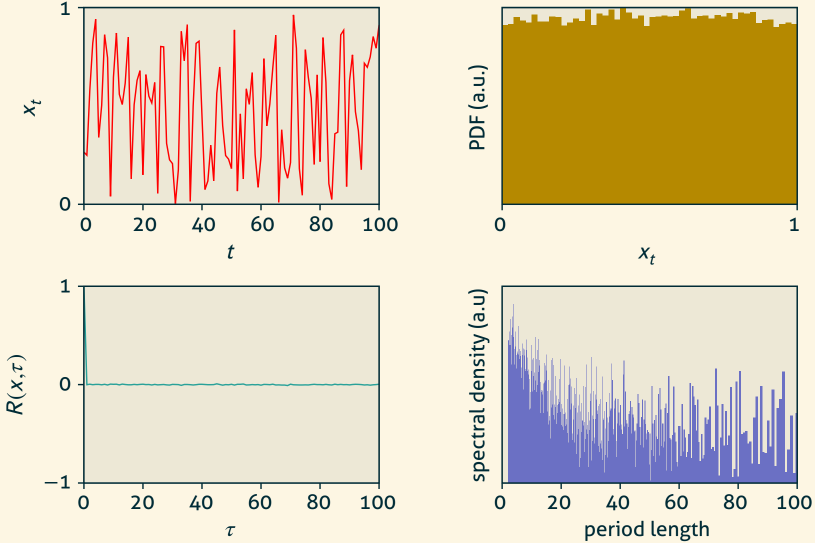
Linear methods can:

- detect periodic processes
  (non-decaying autocorrelation, discrete Fourier spectrum)
- hint at non-stochastic dynamics
  (not normally distributed)
- yield data-based, linear models
  that may not capture essential dynamical properties

Linear methods cannot:

- robustly destinguish noise from chaos
- yield non-linear or chaotic models

Zaslavskii Map



Zaslavskii Map

$$x_{t+1} = \left(x_t + \nu(1 + \mu y_t) + \varepsilon \nu \mu \cos(2\pi x_t)\right) \bmod 1$$
$$y_{t+1} = \left(y_t + \varepsilon \cos(2\pi x_t)\right) \exp(-\Gamma)$$

$\Gamma = 3, \mu = \frac{1-\exp(-\Gamma)}{\Gamma}, \nu = \frac{400}{3}, \varepsilon = 0.3$

Stochasticity vs. Deterministic Chaos

- Simple chaotic maps may be indistinguishable from stochastic processes with linear methods.
- Any pseudo-random-number generator is nothing but a very complex chaotic map.
- But: Nature may be more benign.

Any sufficiently complex deterministicity
is intdistinugishable from stochasticity