TSA Tutorial 1
28.4.2021

Motivation, outlook and key concepts

Manuel Brenner
manuel.brenner@zi-mannheim.de

Organizational Notes

If you hand in as a group, it’s safer if both of you upload the solutions in case we overlook the second name on your sheet

Please upload the solutions straight on the moodle

Please upload only 1 file (html or pdf), and only the solutions!

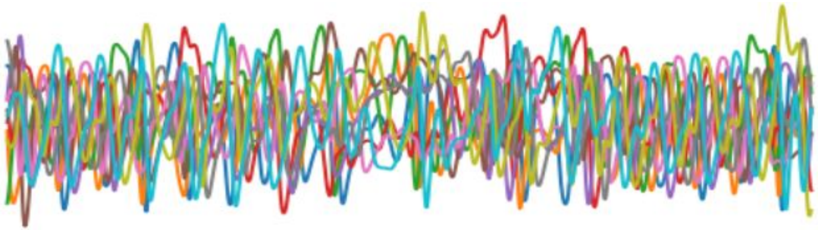
You only need to implement functions yourself when we explicitly ask for it, otherwise feel free to use packages etc.

Time series analysis

Pretty much all real-world systems, such as the brain, are dynamical systems

Most of these systems are generally nonlinear and exhibit multi-scale behavior in both space and time

How well can we understand these systems with a data-driven approach?



Modern Dynamical Systems+Time Series Analysis

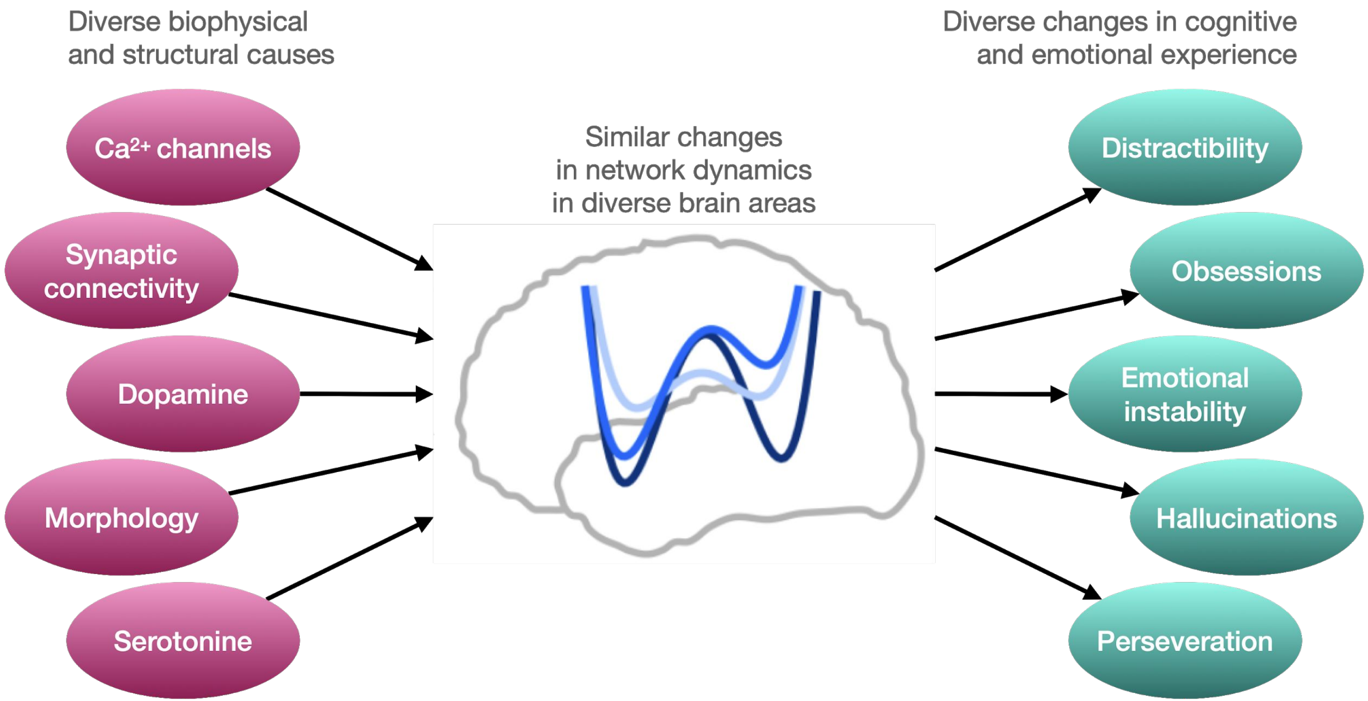
1. **Future state prediction**: e.g. in meteorology and climatology, we seek predictions of the future state of a system

2. **Interpretability and physical understanding**. Provide physical insight and interpretability into a system’s behavior through analyzing trajectories and solutions to the governing equations of motion, e.g. models of cognition

3. **Design and treatment.** How can we tune the parameters of a system for improved performance or stability, e.g. medication in mental illness?

4. **State Estimation:** How can we estimate the full state of the system from limited measurements?

Motivation


Diverse biophysical and structural causes
Ca^{2+} channels
Synaptic connectivity
Dopamine
Morphology
Serotonine
Similar changes in network dynamics in diverse brain areas
Diverse changes in cognitive and emotional experience
Distractibility
Obsessions
Emotional instability
Hallucinations
Perseveration

Dreaming Computers

Spontaneously generated content

Following a meaningful probability distribution

Generated without external inputs

Reflecting content from waking reality

Generative Models+Unsupervised Learning

“What I cannot create, I do not understand.”
Richard Feynman

Find meaningful structure within fluctuating data

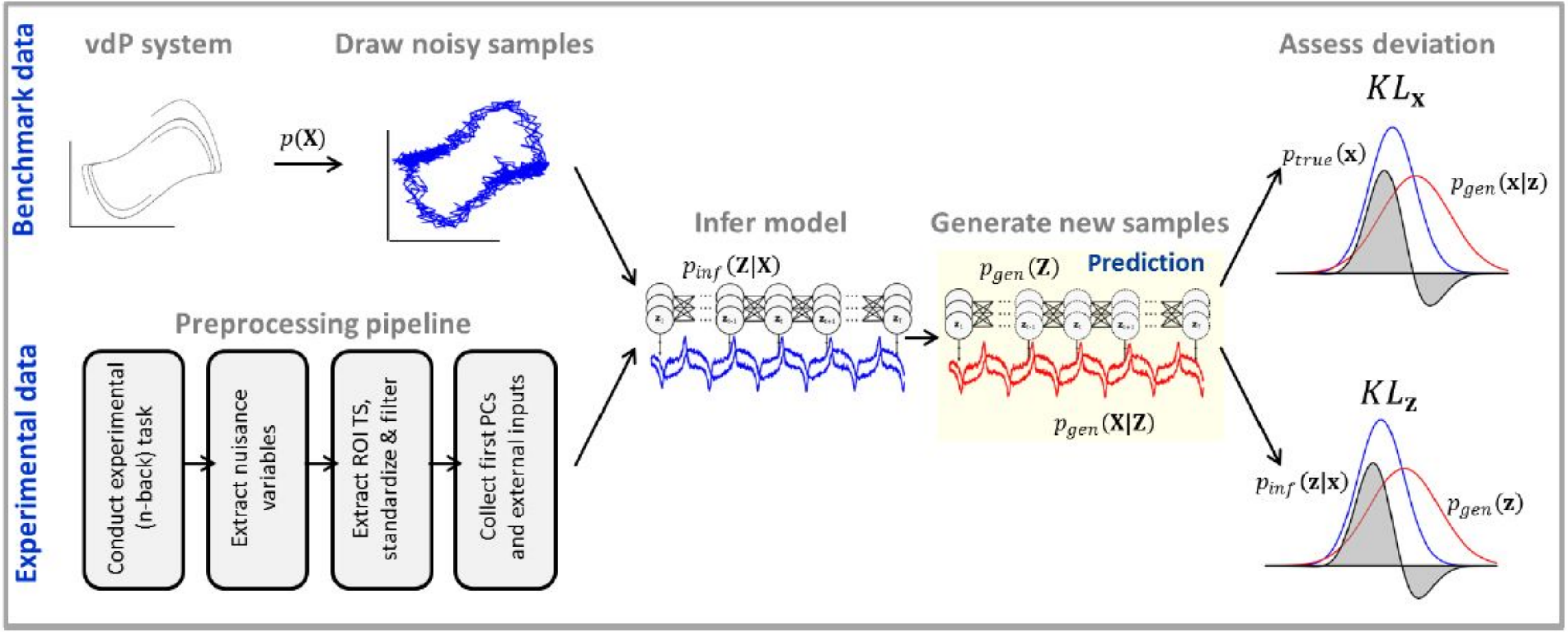
Assume that data is not fully random, but *generated by some process* whose (lower-dimensional) structure we can deduce

We don’t need labels, but simply the ability to predict new data → invert and improve our model based on the quality of the prediction

Different shades of randomness

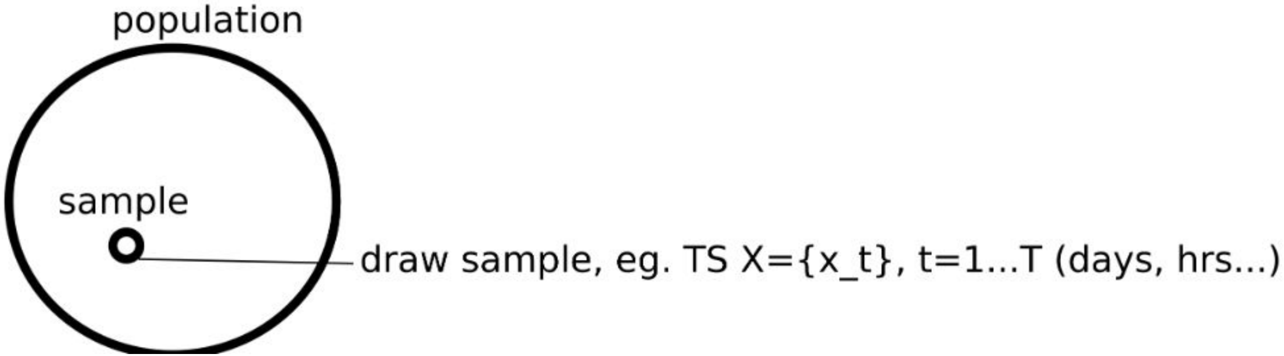
Complete Randomness vs. StyleGAN

Inferring models from data



Statistics Recap

We want to describe a population and that this population (or the time series of this population) is generated by a “true distribution”



→ we are drawing a sample from ***this population of infinite time series*** from which we then are estimating the underlying distribution through statistical estimators

Statistics Recap

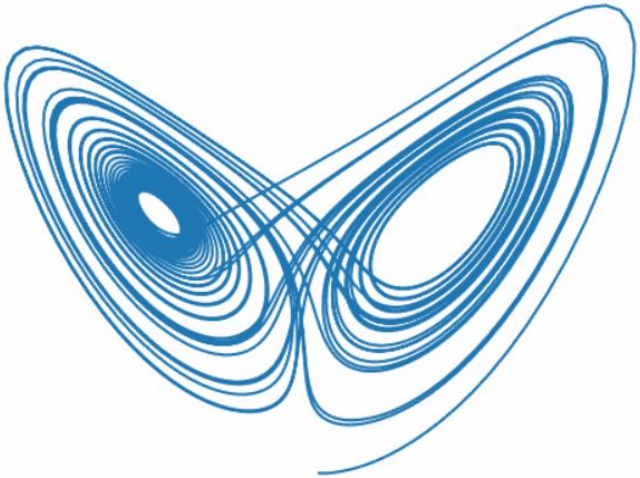
Underlying distribution can be characterized by a dynamical system, e.g. a Lorenz system, gives us a distribution in a phase space (“attractor”)

We can potentially sample ***infinitely many points*** from this distribution by running the ODEs:

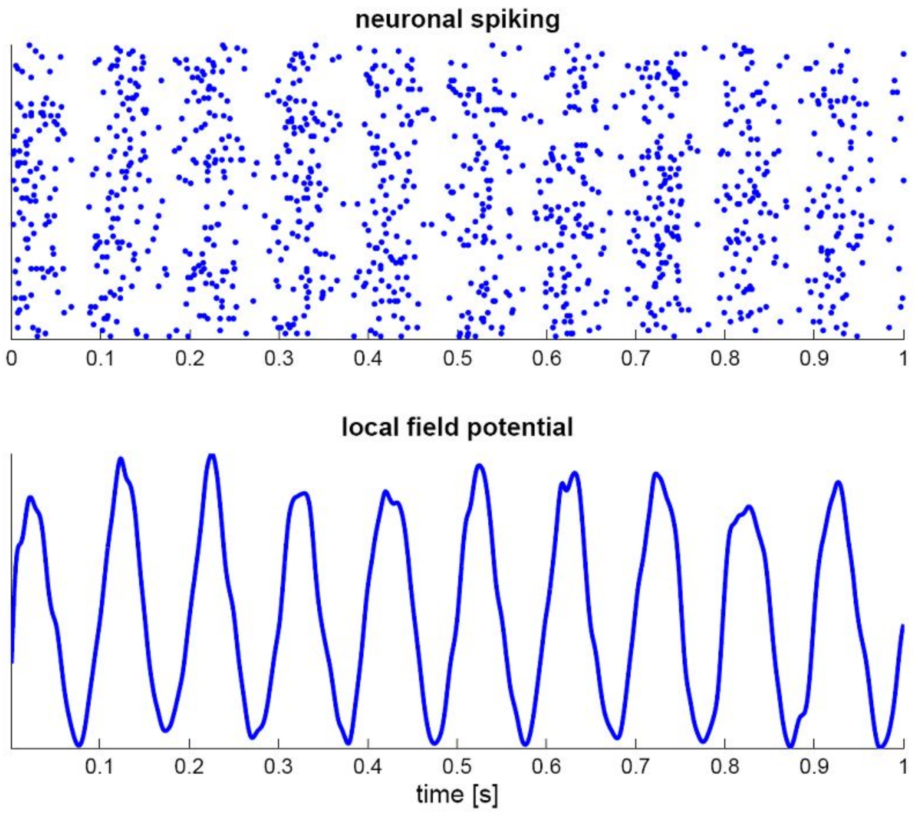
$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x),$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y,$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.$$



Underlying Processes



Statistics Recap

A procedure that is at the backbone of what a lot of researchers do is model estimation through ***parameterized models,*** e.g. AR models, Recurrent Neural Networks, Autoencoders, GANs...:

$$z_t = \boldsymbol{A} z_{t-1} + \boldsymbol{W}\phi(z_{t-1}) + \boldsymbol{h}_0 + \boldsymbol{C}\boldsymbol{s}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

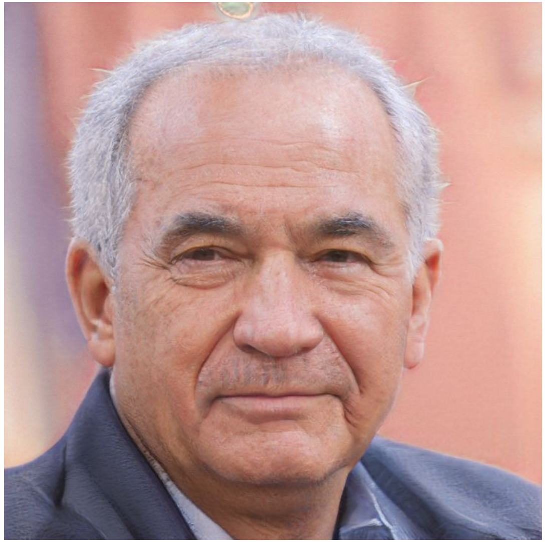
We try to estimate the parameters of the model from the sample in a way as to make it ***capture the whole underlying distribution/data generating process in the best possible way***

Other generative models

E.g. GANs, Variational Autoencoders:

Understand the “true distribution” behind pictures of faces → learn a generative model that allows us to sample completely new faces (e.g. StyleGANs)

→ our aim is to create “deep fakes” of time series data



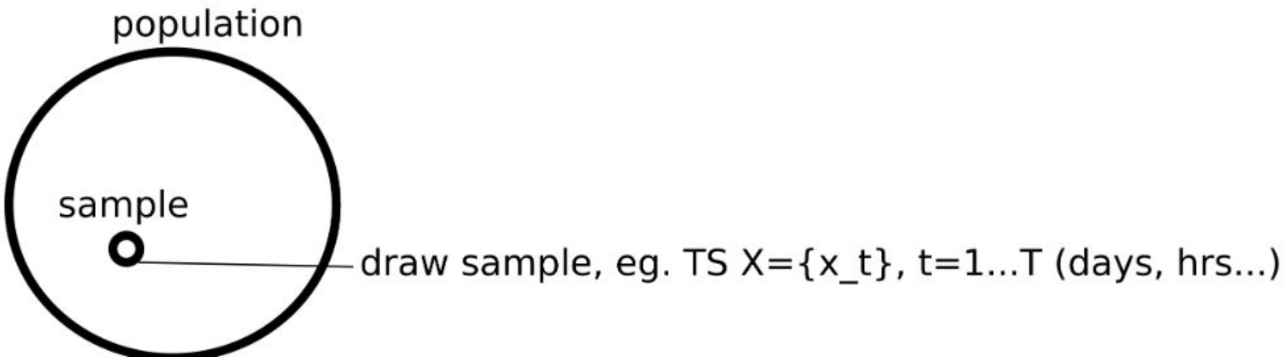
Connections to modern neuroscience theories

- The brain builds internal models of your surroundings, cats, colours, friends, faces...
- Assumption: these models are also *probabilistic* and *generative (e.g. you can close your eyes and dream)*
- Brain updates its models based on its predictions in a statistically meaningful way (***Inference***)
- See f.e. Bayesian Brain Hypothesis, Predictive Coding, Free Energy Principle

---> interplay between machine learning and cognitive science

Statistics Recap

We want to describe a population and that this population (or the timeseries of this population) is generated by a true distribution ”data-gen”



→ we are drawing a sample from this population of TS from which we then are estimating the underlying distribution through statistical estimators

Ergodicity

Key concept:

ergodicity:

$$E_i[x_t^{(i)}] = \frac{lim}{N->\infty} \frac{1}{N} \sum_{i=1}^{N} x_t^{(i)} = E_t[x_t^{(i)}] = \frac{lim}{T->\infty} \frac{1}{T} \sum_{t=1}^{T} x_t^{(i)}.$$

Replace N and t

→ we observe the same behavior averaged over time as averaged over the space of all time series

Stationarity

Properties of the time series do not change across time.

Weak Stationarity:

$$E[x_t] = \mu = const. \quad acov(x_t, x_{t+\Delta t}) = acov(\Delta t) \forall \Delta t$$

Strong stationarity:

$$F(\{x_t|\ t_0 \leq t < t_1\}) = F(\{x_t|\ t_0 + \Delta t \leq t < t_1 + \Delta t\}) \forall\ t_0,\ t_1, and\ \Delta t.$$

→ Assume that we have access to a large sample of timeseries generated by the same underlying process from which we can take ***expectancies across all series i at time t to evaluate the first moments***.

Ergodicity+Stationarity

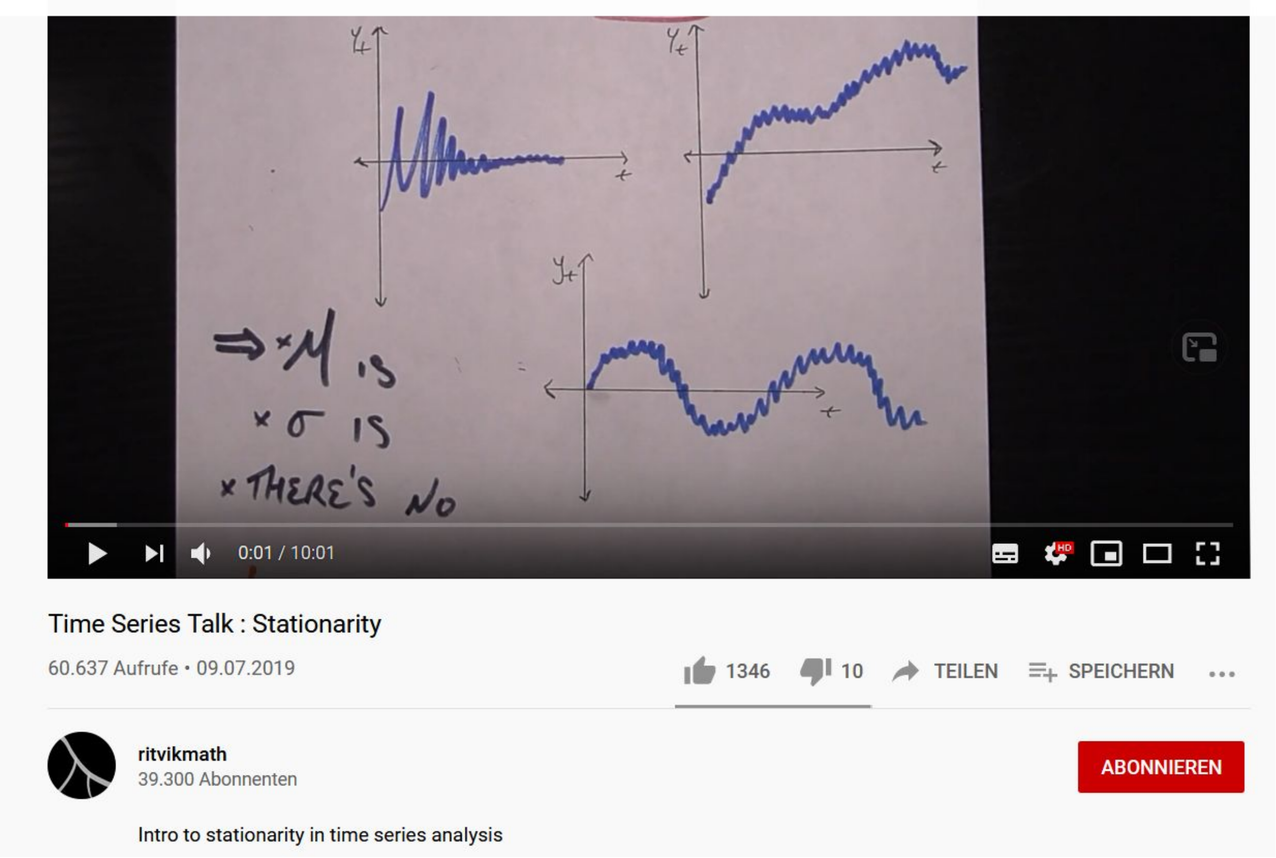
In practice, it may be difficult to determine stationarity empirically in the strict sense ***since we only have access to one or a couple of time series (***see exercise sheet 1, neural measurements etc.)

→ we employ ergodicity to estimate stationarity

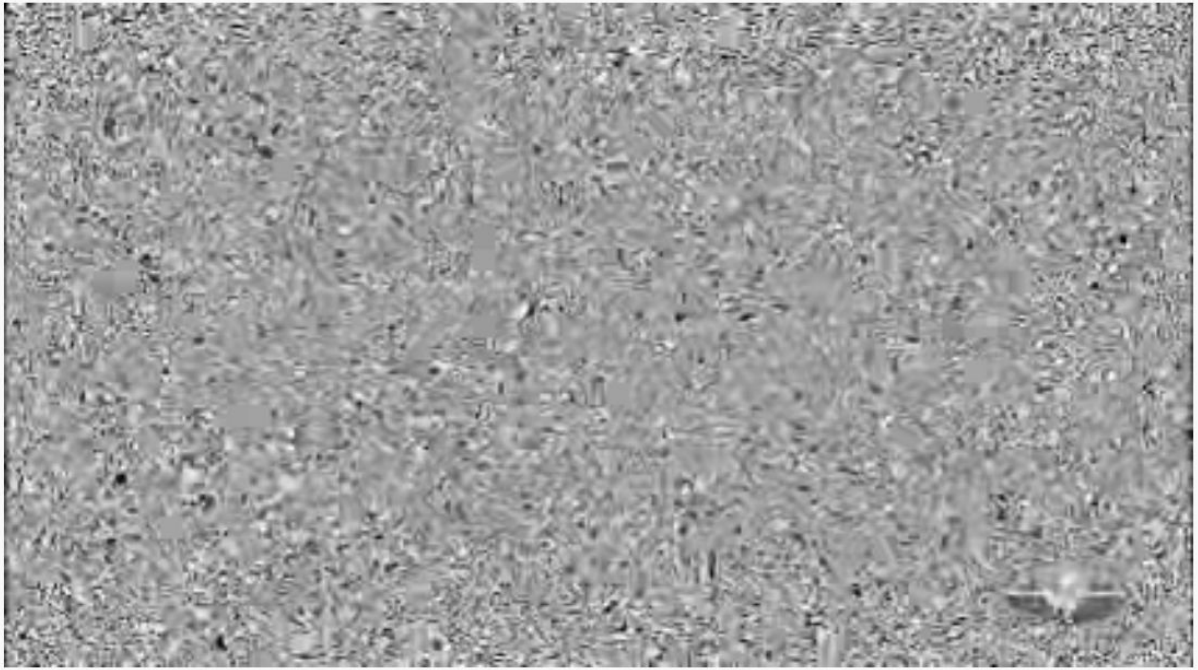
We determine stationarity by

- eyeballing the curve
- Augmented Dickey-Fuller (ADF) gives a number (p-value)

Further Material



White Noise



#whitenoise #tenhournoise #dalesnale

White Noise Ten Hours - Ambient Sound - Masker

2.901.588 Aufrufe • 02.11.2012 10.687 1334 TEILEN SPEICHERN

White Noise

$$E[x_t] = 0 \qquad E[x_t, x_t'] = \sigma^2 for\ t = t', \quad and\ 0\ otherwise,$$

We often assume models decompose into deterministic + noise part

We hope that the noise is WN because it makes everything easier

We can ***always decompose*** our signal according to a deterministic part and a noise part

Why is it called white noise? Is white noise stationary?

Wiener Khintchin theorem: There’s a 1-1 relationship between the acorr function and the power spectrum provided the time series is weak-sense stationary, meaning if you know one you know the other

Linear models vs. nonlinear models

- Nonlinearity remains a primary challenge in analyzing and controlling dynamical systems
- Linear systems are ***completely characterized*** in terms of the spectral decomposition (i.e., eigenvalues and eigenvectors) of the matrix A

  ---> general procedures for prediction, estimation,and control

- No overarching framework exists for nonlinear systems

  → developing a general framework is a mathematical grand challenge of the 21st century.

Simple first step: start with a linear model

ARMA models

We assume we have correlation between events: use these correlations by implementing linear lag operator

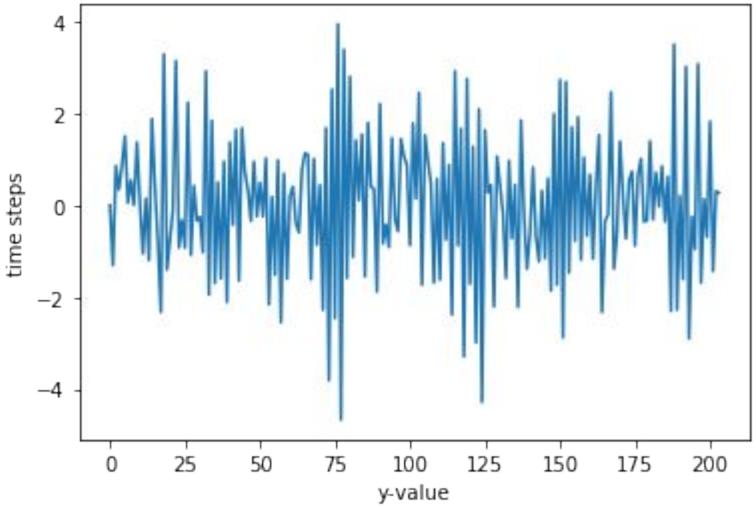
→ autoregressive (AR) and autoregressive moving average (ARMA)

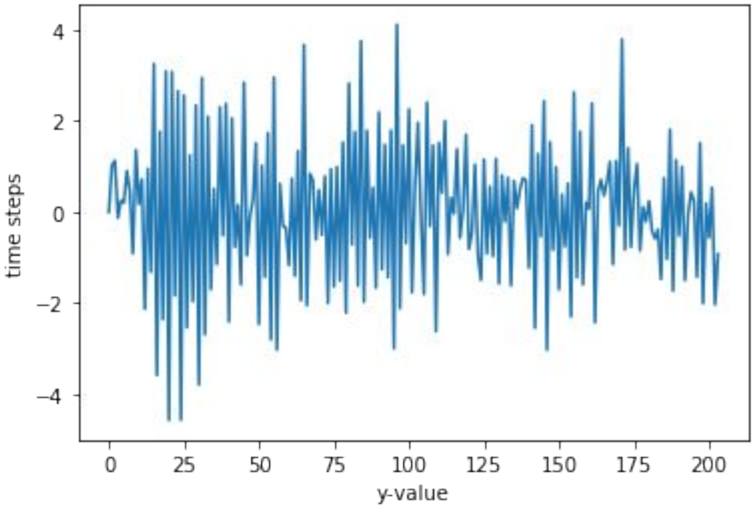
$$x_t = a_0 + \textstyle\sum_{i=1}^{p} a_i x_{t-i} + \sum_{j=0}^{q} b_j \epsilon_{t-j}, \quad \epsilon_t \; W(0, \sigma^2)$$

ARMA models are our first statistical models of our time series that form a ***generative process***

ARMA models

ARMA models are generative processes → we can potentially sample an infinite amount of time series from an initial state x_0





Duality of AR and MA part

We can express models as either pure AR or pure MA processes, by infinitely expanding it as the function of the other

$$x_t = a_0 + a_1 x_{t-1} + \epsilon_t := AR(1)$$

$$\begin{aligned}
x_t &= a_0 + a_1 x_{t-1} + \epsilon_t \\
&= a_0 + a_1(a_0 + a_1 x_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= a_0 + a_1(a_0 + a_1(a_0 + a_1 x_{t-3} + \epsilon_{t-2}) + \epsilon_{t-1}) + \epsilon_t \\
&= a_0 \sum_{i=1}^{t-1} a_1^i + \sum_{i=0}^{t-1} a_1^i \epsilon_{t-i}
\end{aligned}$$

Convergence/Divergence of ARMA models

Expectation at time t is defined by geometric series

$$E[x_t] = E[a_0 \sum_{i=1}^{t-1} a_1^i] + E[\sum_{i=0}^{t-1} a_1^i \epsilon_{t-i}] = a_0 \sum_{i=1}^{t-1} a_1^i + 0 = a_0 \sum_{i=1}^{t-1} a_1^i \text{ (geometric series).}$$

converges to $a_0 \sum_{i=1}^{t-1} a_1^i \to \frac{a_0}{1-a_1}$ if and only if $|a_1| < 0$.

→ see exercise 3)