Friedrich-Alexander-Universität
Technische Fakultät

Machine Learning
Data Analytics

FAU

Machine Learning for Time Series

(MLTS or MLTS-Deluxe Lectures)

Dr. Dario Zanca

Machine Learning and Data Analytics (MaD) Lab
Friedrich-Alexander-Universität Erlangen-Nürnberg
18.10.2022

Organisational Information

Machine Learning for time series

- 5 ECTS
- Lectures + Exercises

Machine Learning for Time Series (Deluxe)

- 7.5 ECTS
- Lectures + Exercises + Project

Topics overview

- Time series fundamentals and definitions (2 lectures) ←
- Bayesian Inference (1 lecture)
- Gussian processes (2 lectures)
- State space models (2 lectures)
- Autoregressive models (1 lecture)
- Data mining on time series (1 lecture)
- Deep learning on time series (4 lectures)
- Domain adaptation (1 lecture)

Course times

Lectures (online)

A new lecture recording is generally released every Tuesday

Consultation hours starting on November 8th, h. 8:15 – 9:30

Exercises (online)

Live Zoom Session starting on November 9th

Recordings will be uploaded

Project (online)

Introduction during the first exercise Live Zoom Session (November 9th)

Exams and evaluation

Written Exam (5 ECTS)

- Likely written or online
- 70% from lectures, 30% from exercises

Project (2.5 ECTS) - Optional

- Define your own project topics
- Work in teams of 2-3 students
- Independent work
- Write a scientific report
- Peer-review your co-student’s reports

Course organizers
Lecturers

Machine Learning and Data Analytics (MaD) Lab

- Dr. Dario Zanca, dario.zanca@fau.de *
- Prof. Dr. Björn Eskofier, bjoern.eskofier@fau.de

* Please, address all your correspondence about the course to Dr. Dario Zanca

Course organizers
Teaching assistants

Exercises, responsibles:

- Richard Dirauf (M.Sc.), richard.dirauf@fau.de
- Philipp Schlieper (M.Sc.), philipp.schlieper@fau.de

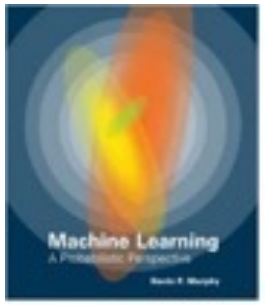
Projects, responsibles:

- Dr. Dario Zanca, dario.zanca@fau.de *
- Johannes Roider (M.Sc.), johannes.roider@fau.de

References

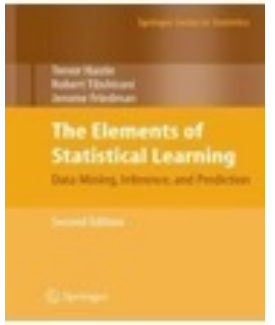
Machine learning: A Probabilistic Perspective,

by Kevin Murphy (2012)



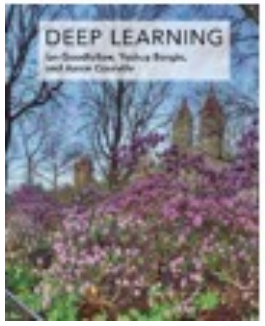
The Elements of Statistical Learning: Data Mining, Inference, and Prediction

by Trevor Hastie, Robert Tibshirani, and Jerome Friedman (2009)



Deep Learning

by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016)



Time series fundamentals

Motivations

An old history of time series analysis: Babylonian astronomical diaries

VII century B.C.

"[…] Night of the 5th, beginning of the night, the moon was 2 ½ cubits behind Leonis […] Night of the 17th, last part of the night, the moon stood 1 ½ cubits behind Mars, Venus was below."

- Babylonians collected the earliest evidence of periodic planetary phenomena
- Applied their mathematics for systematic astronomic predictions

An old history of time series analysis: Babylonian astronomical diaries

Nowadays, thousands of ground-based and space-based telescopes[(a)] generate new knowledge every night.

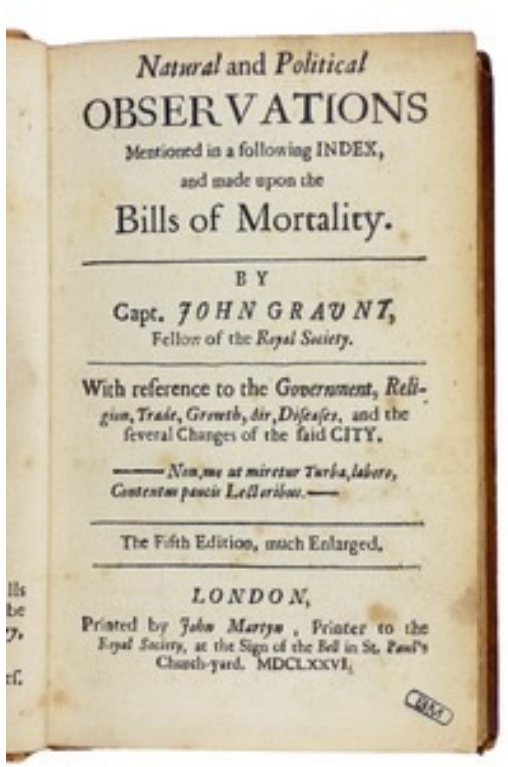
- The Vera C. Rubin Observatory in Chile is geared up to collect 20 terabytes per night from 2022[(b)].
- The Square Kilometre Array, the world's largest radio telescope, will generate up to 2 petabytes daily, starting in 2028.
- The Very Large Array (ngVLA) will generate hundreds of petabytes annually.

[(a)] https://research.arizona.edu/stories/space-versus-ground-telescopes
[(b)] https://www.nature.com/articles/d41586-020-02284-7

An old history of time series analysis: The Birth of Epidemiology

1662, John Graunt describes the data collection:

"When anyone dies, […] the same is known to the Searchers, corresponding with the said Sexton. The Searchers hereupon...examine by what Disease, or Casualty the corps died. Hereupon they make their Report to the Parish-Clerk, and he, every Tuesday night, carries in an Accompt of all the Burials, and Christnings, hapning that Week, to the Clerk of the Hall."


Natural and Political
OBSERVATIONS
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.
BY
Capt. JOHN GRAUNT,
Fellow of the Royal Society.
With reference to the Government, Religion, Trade, Growth, Air, Diseases, and the several Changes of the said CITY.
The Fifth Edition, much Enlarged.
LONDON,
Printed by John Martyn, Printer to the Royal Society, at the Sign of the Bell in St. Paul's Church-yard. MDCLXXVI.

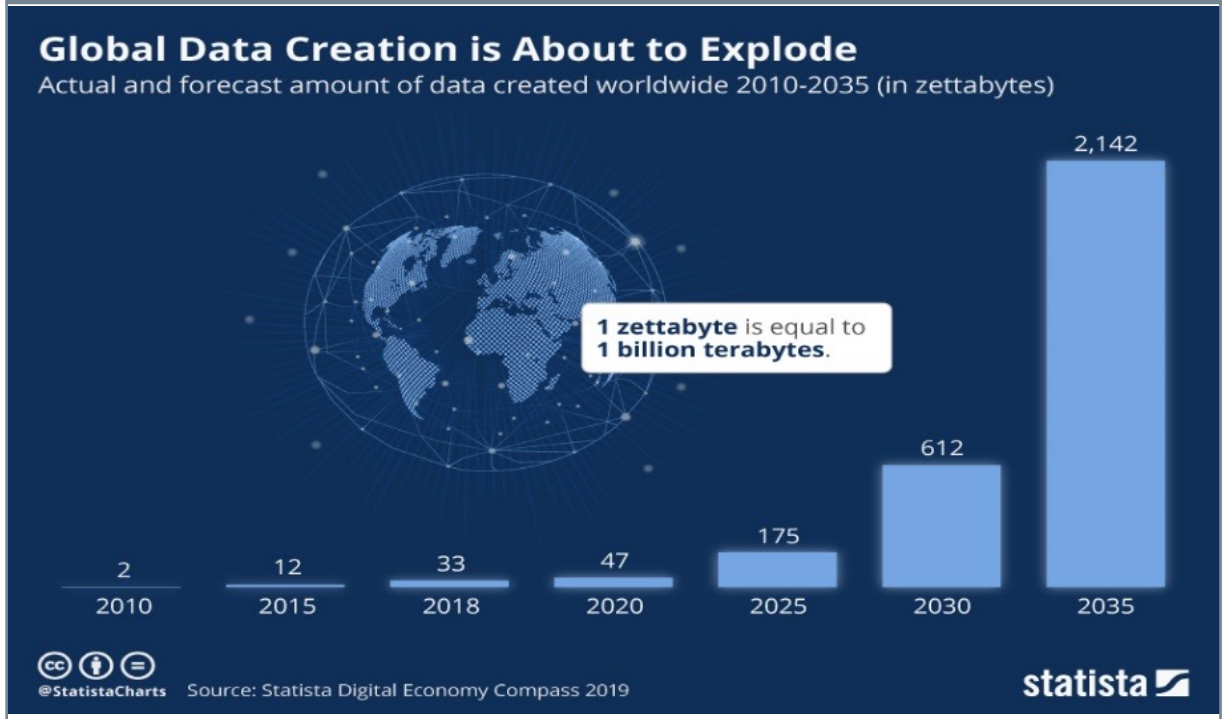
- Rudimentary conclusions about the mortality and morbidity of certain diseases
- Graunt's work is still used today to study population trends and mortality

Importance of time series

Machine learning on time series is becoming increasingly important because of the massive production of time series data from diverse sources, e.g.,

- Digitalization in healthcare
- Internet of things
- Smart cities
- Process monitoring

The amount of created data increased from two zettabytes in 2010 to 47 zettabytes in 2020



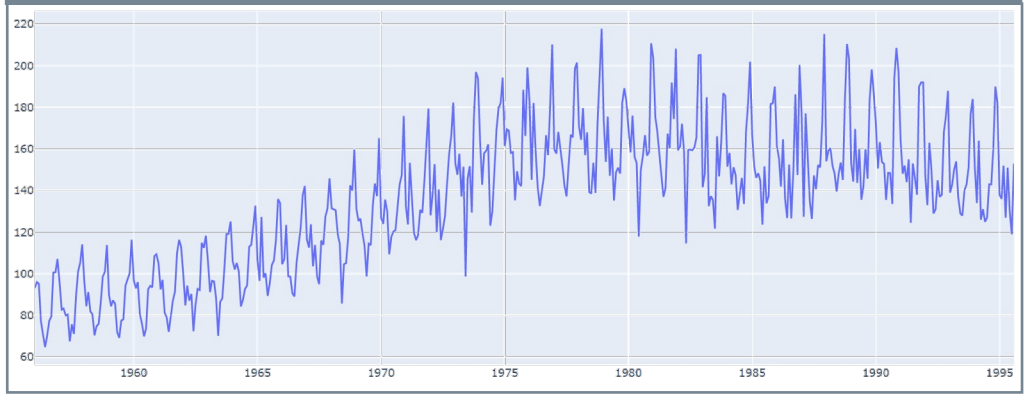
https://www.statista.com

Example: Predicting demand of amazon products

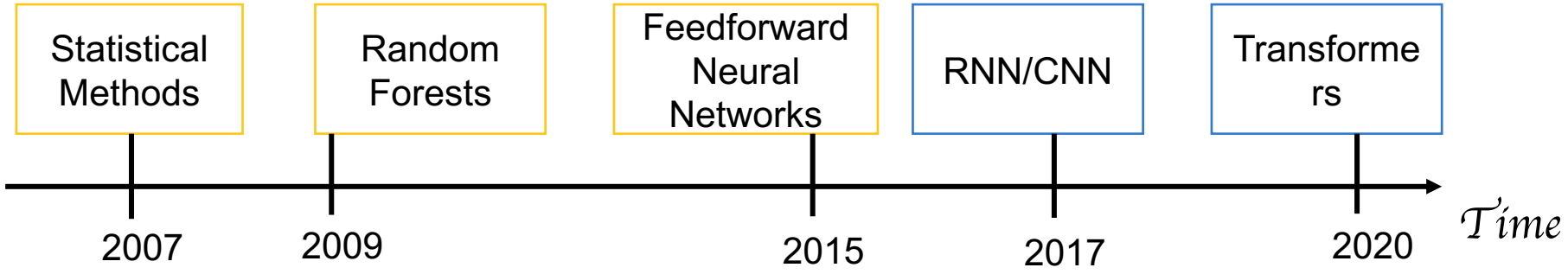
Amazon sells 400 million products in over 185 countries[(a)].

- Maintaining surplus inventory levels for every product is cost-prohibitive.
- Predict future demand of products

Some products are sold depending on the season



Methods:

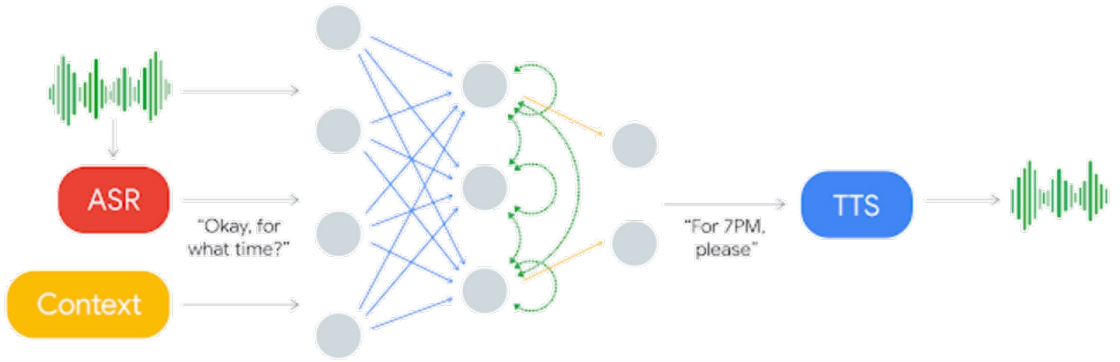


□ First models required manual feature engineering

□ New methods are fully data-driven

Example: Google Duplex makes tedious phone calls

Method: An RNNs with several features. We use a combination of text to speech (TTS) engine and a synthesis TTS engine to control intonation (e.g., “hmm”s and “uh”s).



Limitations: trained on specific tasks. Cannot deal general conversations.

- Additional audio features
- Automatic speech recognition
- Desired service, time/day

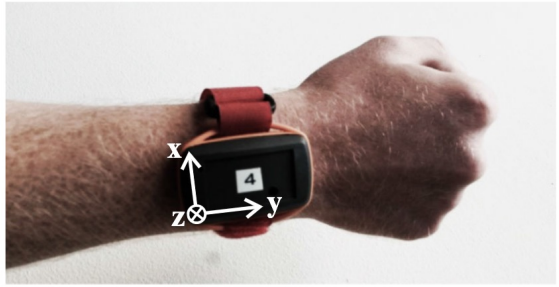
E.g., Duplex calling a restaurant.

https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

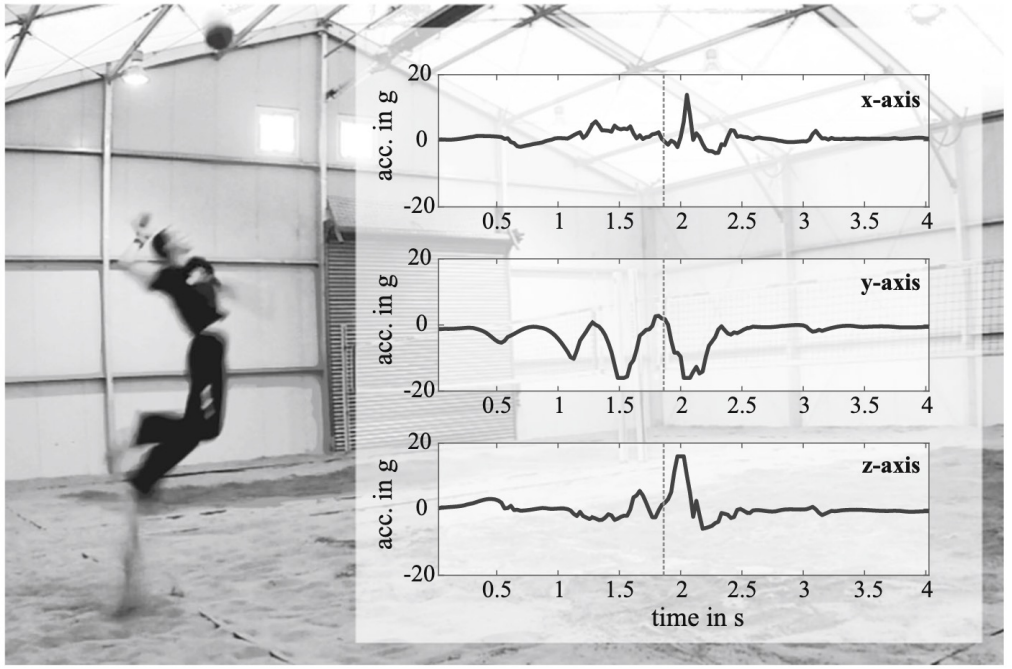
Example: Activity recognition in sports (FAU Erlangen)

Many injuries in sports are caused by overuse.

- These injuries are a major cause for reduced performance of professional and non-professional beach volleyball players.
- Monitoring of player actions could help identifying and understanding risk factors and prevent such injuries.



Sensor attachment at thewrist of the dominant hand witha soft, thin wristband



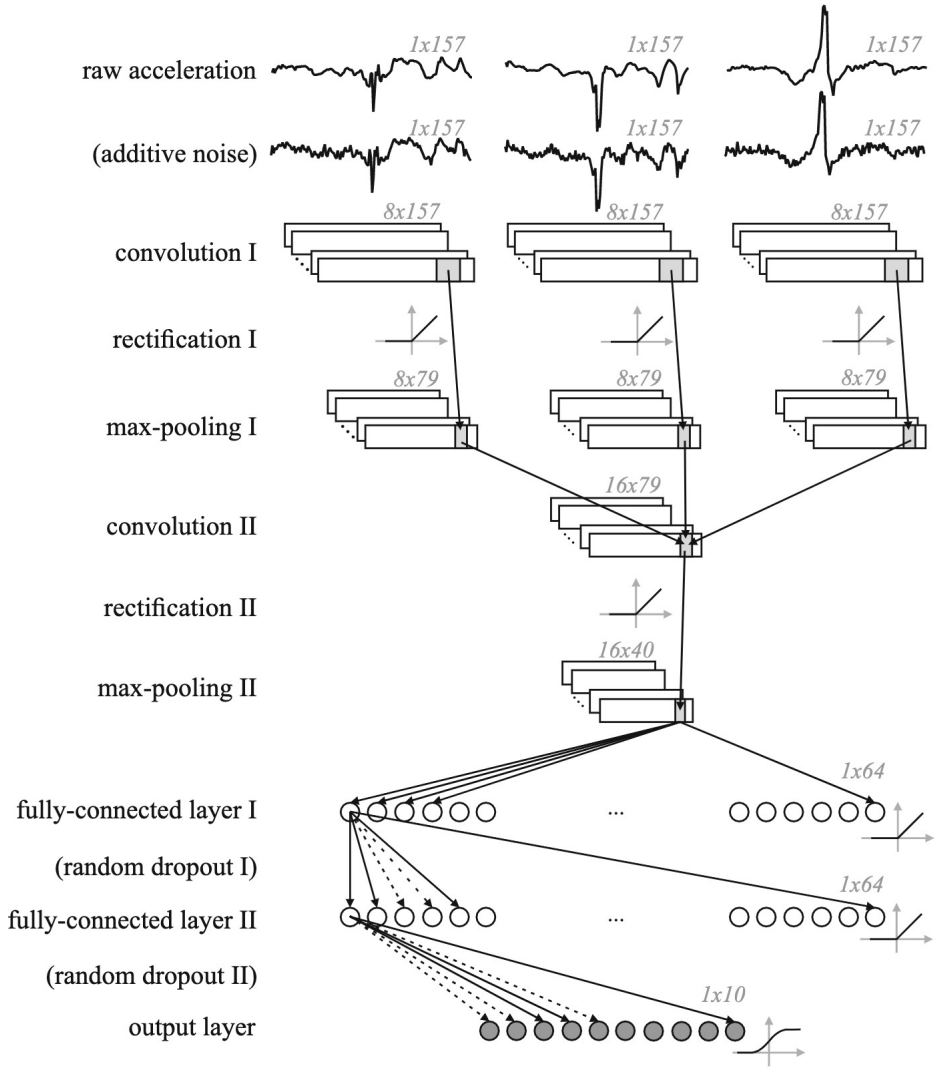
https://doi.org/10.1007/s10618-017-0495-0

Example: Activity recognition in sports (FAU Erlangen)

Method: A CNN is used to classify players' activities. Classifications allow to create players' profiles.

Actions:

- Underhand serve
- Overhand serve
- Jump serve
- Underarm set
- Overhead set
- Shot attack
- Spike
- Block
- Dig
- Null class.



https://doi.org/10.1007/s10618-017-0495-0

Time series fundamentals

Definitions and basic properties

What is a time series?

A time series can be described as a set of observations, taken sequentially in time,

$$S = \{s_1, \dots, s_T\}$$

where $s_i \in \mathbb{R}^d$ is the measured state of the observed process at time t_i.

Typically, observations are generally *dependent*

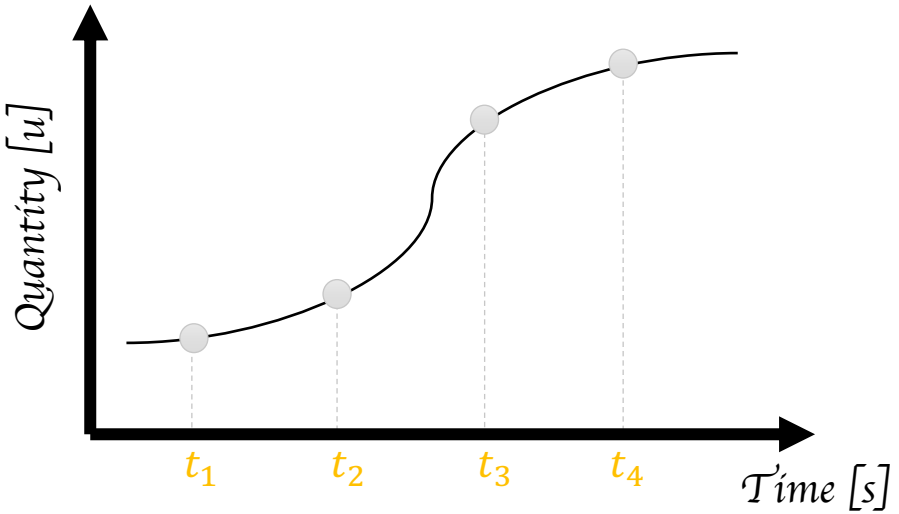
- Studying the nature of this dependency is of particular interest
- Time series analysis is concerned with techniques for the analysis of these dependencies

Terminology: Regularly Sampled vs Irregularly Sampled

Discrete time series are **regularly sampled** if their observations are euqally spaced in time.

$$\forall i \in \{1, \dots, T-1\},$$

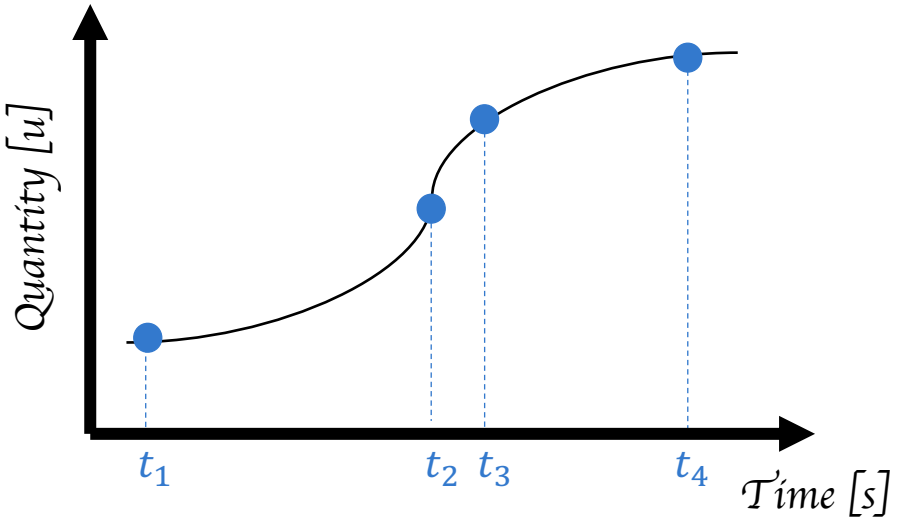
$$\Delta_{t_i} = t_{i+1} - t_i = const.$$



In contrast, for **irregularly sampled** time sequences, the observations are not equally spaced.

- They are generally defined as a collection of pairs

$$S = \{(s_1, t_1), \dots, (s_T, t_T)\}$$



Terminology: Univariate vs Multivariate

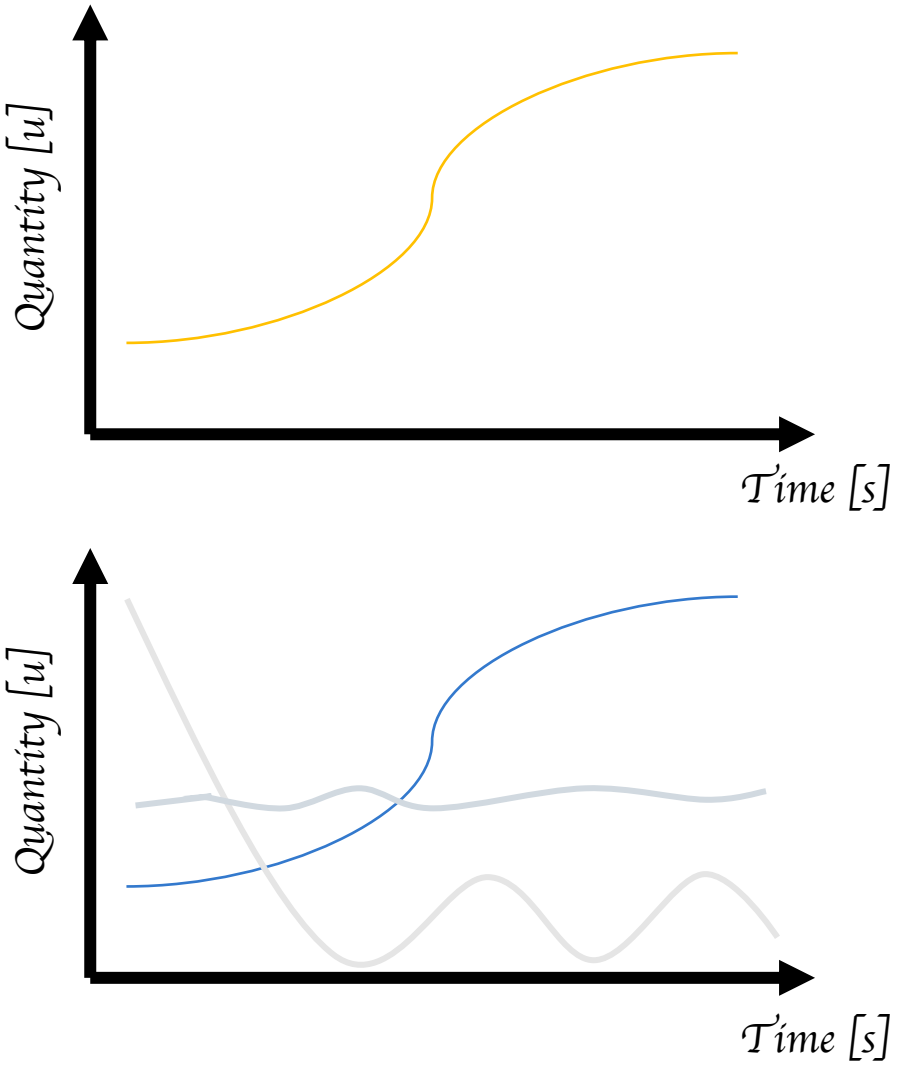
Let $S = (s_1, \dots, s_T)$ be a time series, where $s_i \in \mathbb{R}^d, \forall i \in \{1, \dots, T\}$.

If $d = 1$, S is said **univariate**.

- Only one variable is varying over time.

If $d > 1$, S is said **multivariate**.

- Multiple variables are varying over time
    - E.g., tri-axial accelerometer measurements



Terminology: Discrete vs Continuous

A time series is said to be **continuous** if observations are made at each instant of time, even when its measurements consist only of a discrete set of values.

- E.g., the number of people in a room.

A time series is said to be **discrete** if observations are taken at specific times. Discrete time series can arise in different ways:

- Sampled (e.g., daily rainfall)
- Aggregated (e.g., monthly reports of daily rainfalls)

Terminology: Discrete vs Continuous

We will denote as **mixed-type** a multivariate time series consisting of both continuous and discrete observations

- E.g., a time series consisting of continuous sensor values and discrete event log for the monitoring of an industrial machine

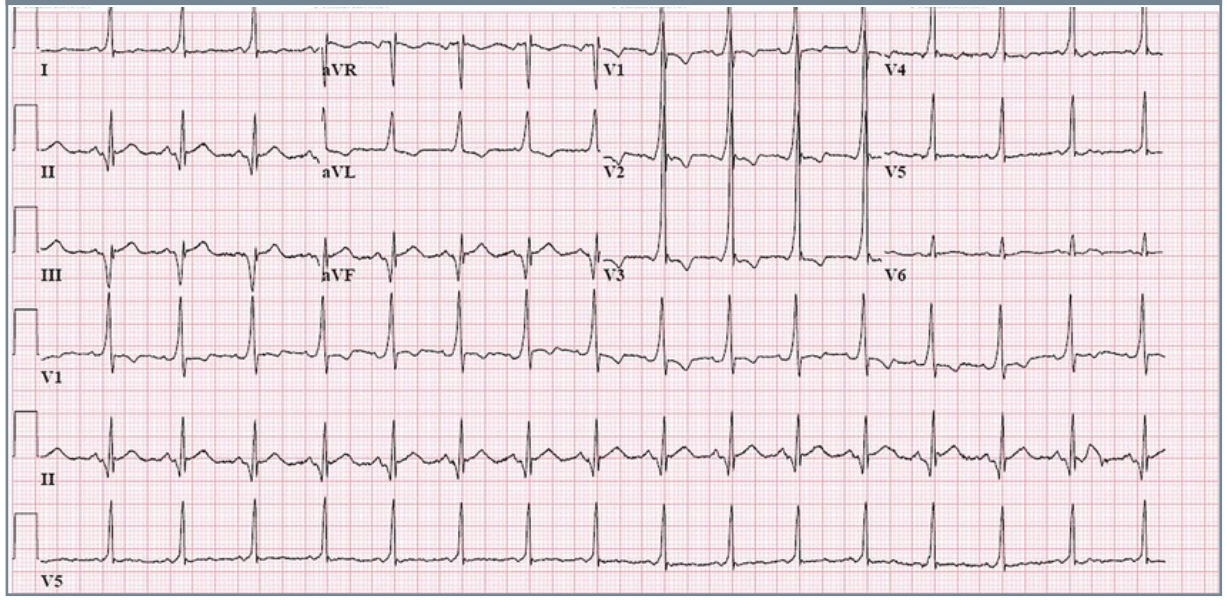
Terminology: Periodic

A time series is said **periodic** if there exists a number $\tau \in \mathbb{R}$, called *period*, such that

$$s_i = s_{i+\tau}\,, \forall i \in \{1, \ldots, T - \tau\}$$

E.g., the continuous time series defined by the trigonometric function $\mathrm{f(x)} = \sin(x)$

Is the biological signal of an heartbeat a periodic function?



Terminology: Deterministic vs Non-Deterministic

A **deterministic** time series is one that could be expressed explicitly by an analytical expression.

- Observations are generated from a system with no randomness.

In contrast, a **non-deterministic** time series can not be described by an analytic expression. A time series may be non-deterministic because :

- The information necessary to describe the process is not fully observable, or
- The process generating the time series is inherently random

Stochastic Process

Non-deterministic time series can be regarded as manifestations (equiv., realization) of a **stochastic process**, which is defined as a set of random variables $\{X_t\}_{t\in\{1,\ldots,\mathrm{T}\}}$

Even if we were to imagine having observed the process for an infinite period T of time, the infinite sequence

$$S = \{\ldots, s_{t-1}, s_t, s_{t+1,\ldots}\} = \{s_t\}_{t=-\infty}^{+\infty}$$

would still be a single **realization** from that process.

Stochastic Process

Still, if we had a battery of N computers generating series $S^{(1)}$, …, $S^{(N)}$, and considering selecting the observation at time t from each series,

$$\left\{s_t^{(1)}, \ldots, s_t^{(N)}\right\}$$

this would be described as a sample of N realizations of the random variable X_t

This random variable X_t is associated with an **unconditional density**, denoted by

$$f_{X_t}(s_t)$$

- E.g., for the Gaussian white noise process $f_{X_t}(s_t) = \frac{1}{\sqrt{2\,\pi\sigma}}\, e^{\frac{-s_t^2}{2\,\sigma^2}}$

Stochastic Process

The **unconditional mean** is the expectation, provided it exists, of the t-th observation, i.e.,

$$E(X_t) = \int_{-\infty}^{+\infty} s_t \, f_{X_t}(s_t) ds_t = \mu_t$$

Similarly, the **variance** of the random variable X_t is defined as

$$E(X_t - \mu_t)^2 = \int_{-\infty}^{+\infty} (s_t - \mu_t)^2 \, f_{X_t}(s_t) ds_t$$

Stochastic Process

Given any particular realization $S^{(i)}$ of a stochastic process (i.e., a time series), we can define the vector of the $j+1$ most recent observations

$$x_t^{\mathrm{i}} = [s_{t-j}^{(i)}, \dots, s_t^{(i)}]$$

We want to know the probability distribution of this vector x_t^{i} across realizations. We can calculate the ***j*-th autocovariance**

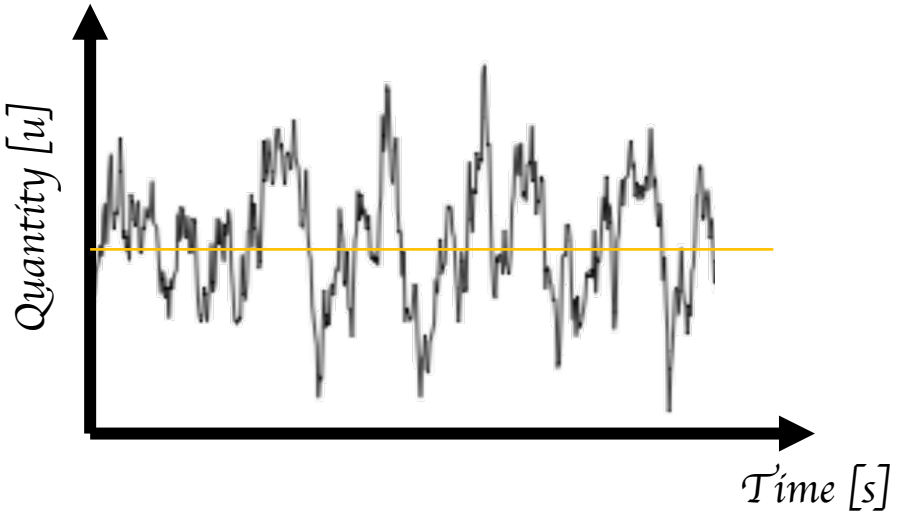
$$\gamma_{jt} = E(X_t - \mu_t)(X_{t-j} - \mu_{t-j})$$

Stationarity

If neither the mean μ_t or the autocovariance γ_{jt} depend on the temporal variable t, then the process is said to be (weakly) **stationary**.

E.g., let the stochastic process $\{X_t\}_{t=-\infty}^{+\infty}$ represent the sum of a constant μ with a Gaussian white noise process $\{\epsilon_t\}_{t=-\infty}^{+\infty}$, such that

$$X_t = \boldsymbol{\mu} + \epsilon_t$$



Then, its mean is constant: $E(X_t) = \mu + E(\epsilon_t) = \mu$

and its j-th autocovariance: $E(X_t - \mu)(X_{t-j} - \mu) = \gamma_j$

In other words: A process is said to be stationary if the process statistics do not depend on time.

Ergodicity

Given a time series, denoted by $S^{(i)} = \left\{s_1^{(i)}, \ldots, s_T^{(i)}\right\}$, we can compute the sample temporal average as

$$\bar{s} = \frac{1}{T}\sum_{t=1}^{T} s_t^{(i)}$$

The ergodicity of a time series bind the concept of the process mean with that of temporal sample mean:

- A process is said to be ergodic if $\bar{s}$ converges to μ_t as $T \to \infty$

In other words: A process is said to be ergodic if its time statistics equals the process statistic, provided that the process is observed long enough.

Example: Stationarity and Ergodicity

To clarify the concept, we give an example of stationary but not ergodic process. Suppose the mean $\mu^{(i)}$ of the i-th realization of $\{X_t\}_{t=-\infty}^{+\infty}$ is sampled from the normal distribution $U(0, \lambda^2)$ and, similarly to the previous example, $X_t^{(i)} = \mu^{(i)} + \epsilon_t$.

We have that the process is stationary because:

$$\mu_t = E\left(\mu^{(i)}\right) + E(\epsilon_t) = 0$$

$$\gamma_{jt} = E\left(\mu^{(i)} + \epsilon_t\right)\left(\mu^{(i)} + \epsilon_{t-j}\right) = \lambda^2$$

Example: Stationarity and Ergodicity

However, its sample temporal mean, converges to a different value than the process mean, i.e.,

$$\bar{s} = (1/T) \sum (\mu^{(i)} + \epsilon_t) = \mu^{(i)}$$

Time series fundamentals

i.i.d. observations and central limit theorems

Time series and i.i.d. data

Observations collected in a time series $S = (s_1, \dots, s_T)$ are **generally not i.i.d.**

- Observation s_i could be **dependent** on previous observations s_j, with $j < i$
- The distribution of the underlying data generation process could change over time, i.e. it is **not identically distributed**

For example:

- The price of a stock today depends on its price yesterday **(dependence)**
- and the volatility of the stock, i.e., its dispersion of returns, might change over time **(change on the underlying distribution)**

Time series and i.i.d. data

The structure of this dependence imposes challenges on the statistical data analysis of time series.

- Many tools for statistical inference are valid only for i.i.d. data

Time series and i.i.d. data

It might be useful to be able to assess the structure of the dependence between random variables. For this reason we make use of their correlation.

- Generally, we measure the correlation between two variables X_i and X_j with their **covariance** $Cov(X_i, X_j)$.
    - $Cov(X_i, X_j) = 0 \rightarrow \text{uncorrelated}$
- We measure dependence of an entire time series with a similar concept, the **long-run variance**
    - $\sigma_i^2 = \sum_{\mathbb{Z}} Cov(X_i, X_{i+h})$

The Central Limit Theorem

The **Central Limit Theorem (CLT)** suggests that the sum of random variables converges to a normal distribution, under precise conditions.

More precisely, for a sequence of i.i.d. random variables $\{X_t\}_{t\in\{1,\dots,T\}}$ with $\mu = E(X_t)$ and $\sigma^2 = E(X_t - \mu)^2$, by the CLT it holds:

$$\sqrt{T}\left(\frac{1}{T}\sum_{1}^{T} X_i - \mu\right) \to \mathcal{N}(0, \sigma^2)$$

For stationary time series with mean μ and long-run variance σ^2 the CLT holds as before.

Why is the CLT important?

If the CLT holds for a time series, we can draw from a larger range of methods.

- Statistical inference depends on the possibility to take a broad view of results from a sample to the population.
- The CLT legitimizes the assumption of normality of the error terms in linear regression.

However,

- Many time series we encounter in the real world satisfy CLT assumption of independence and stationarity
- Or can be transformed into stationary time series, e.g., by differentiations or other transformations

It is a good idea to start by checking whether the data is independent or stationary.

Insight: CLT for dependent random variables

Different version of the CLT exist for dependent random variables. For example, under the assumption of a M-dependent random process[(a)], we have that the following limit theorem holds:

Let $\{X_t\}_{t\in\{1,\dots,T\}}$ be M-dependent stationary process with mean μ, covariance γ_j, and denoted with V_M the variance of the mean of n observations,

$$V_M := \sum_{j=-M}^{M} \gamma_j$$

If $V_M > 0$, then,

$$\sqrt{n}(X_i - \mu) \to N(0, V_M).$$

[(a)] A stochastic process $\{X_t\}_{t\in\{1,\dots,T\}}$ is said to be M-dependent if $\{X_t\}_{t\leq k}$ are independent of the stochastic variables $\{X_t\}_{t\geq k+M+1}$

Time series fundamentals

Recap

Recap

Time series have long been studied in history

- Recent digitalization increases the importance of time series analysis

Properties of time series

- Regularly vs irregularly sampled
- Univariate vs multivariate
- Discrete vs continuous
- Periodic
- Deterministic vs non-deterministic
- Stationarity
- Ergodicity

Central limit theorem only holds for stationary time series

- Less restrictive CLT versions exist
- Need to properly learn dependences