

DS-GA 3001.001
Probabilistic time series analysis
Lecture 1
Logistics. Introduction to time series

Instructor: Cristina Savin
NYU, CNS & CDS

Course logistics

Instructor

Cristina Savin, csavin@nyu.edu

Office hours: TBD

TAs:



Colin
Bredenberg



Jiyuan
Lu



Ashwin
Siripurapu



Artie
Shen

Office hours: TBD

Course page: <https://github.com/savinteachingorg/pTSAfall2020>

Piazza: <https://piazza.com/nyu/fall2020/dsga3001001/home>

Course logistics: grading

5 problem sets 25%

Primarily derivations, 2w for each

Lab work 20%

Coding: **python**, weekly, ideally finished during lab, due end of the week.

Midterm 20%

Oct 27th, ARIMA+ LDS+HMMs

Project 25%

Groups of 2-3, topic of choice

Project proposal due **Oct. 9th**

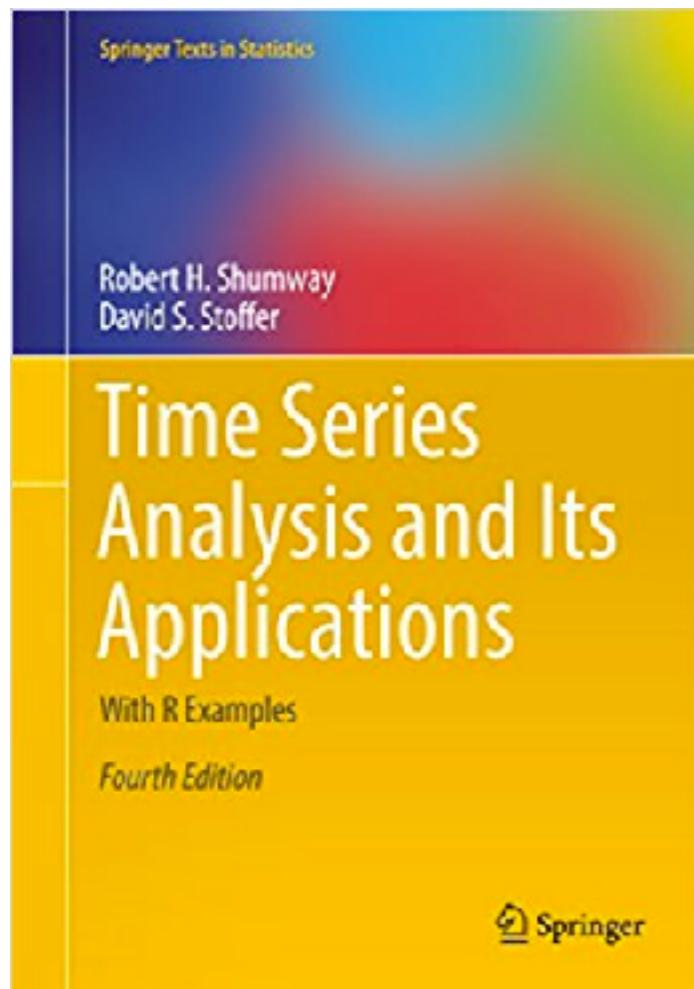
Participation 10%

Class discussions, office hours, piazza

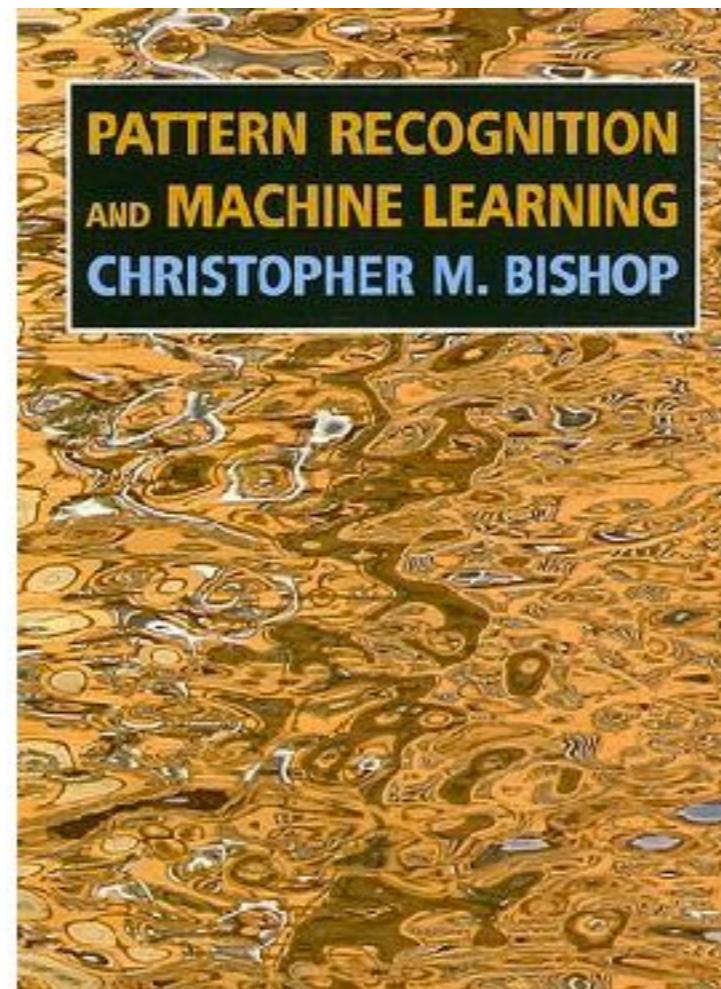
No official prerequisites, but if you haven't seen much **probability** and **linear algebra** recently,
a quick refresher is strongly encouraged

Bibliography

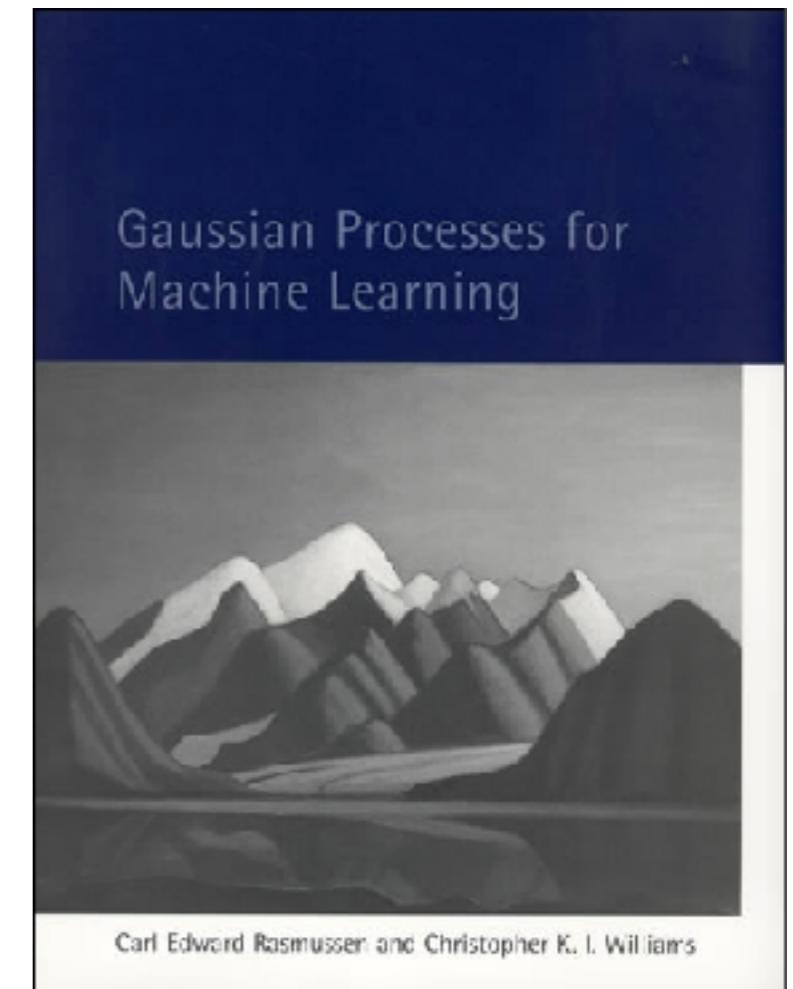
No course textbook, handouts for each section. Lectures based on:



**Intro, AR(I)MA,
Spectral methods**



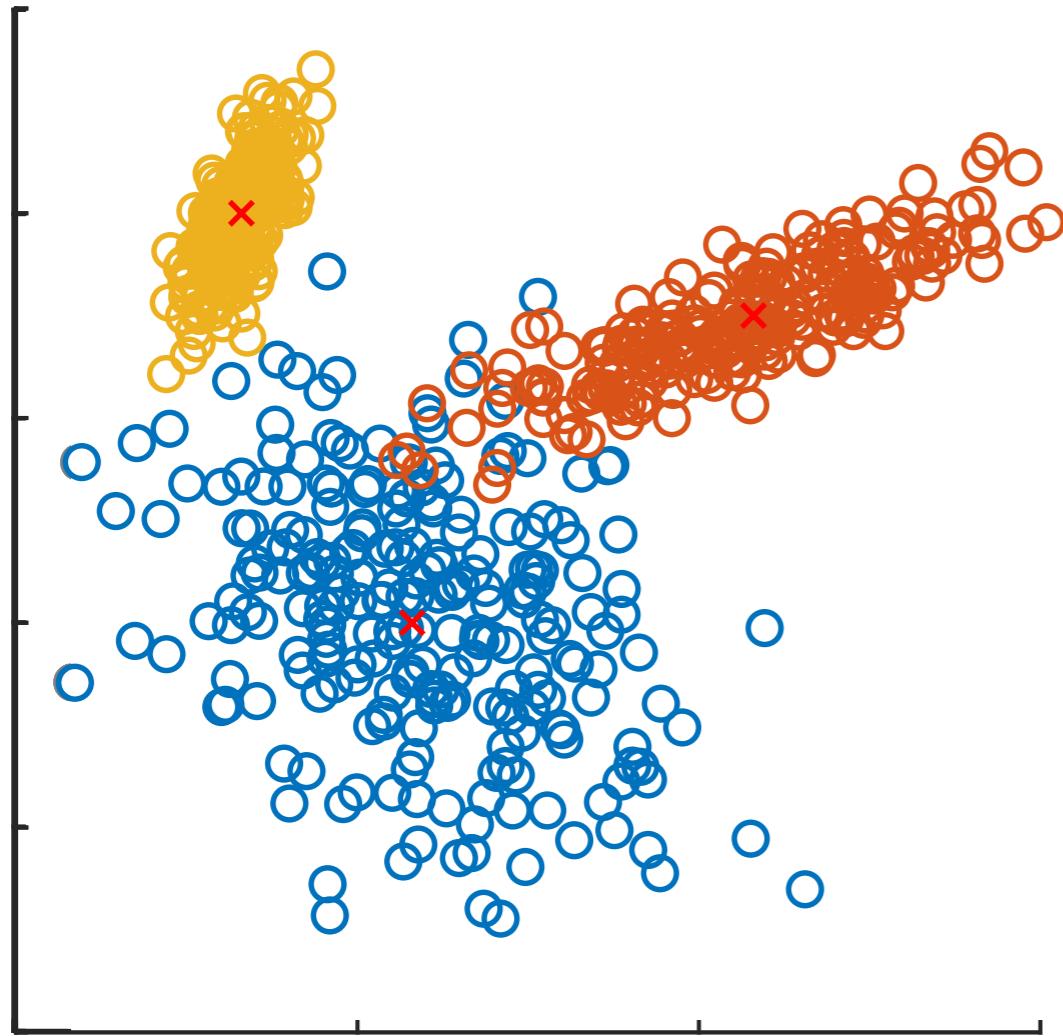
Latent space models



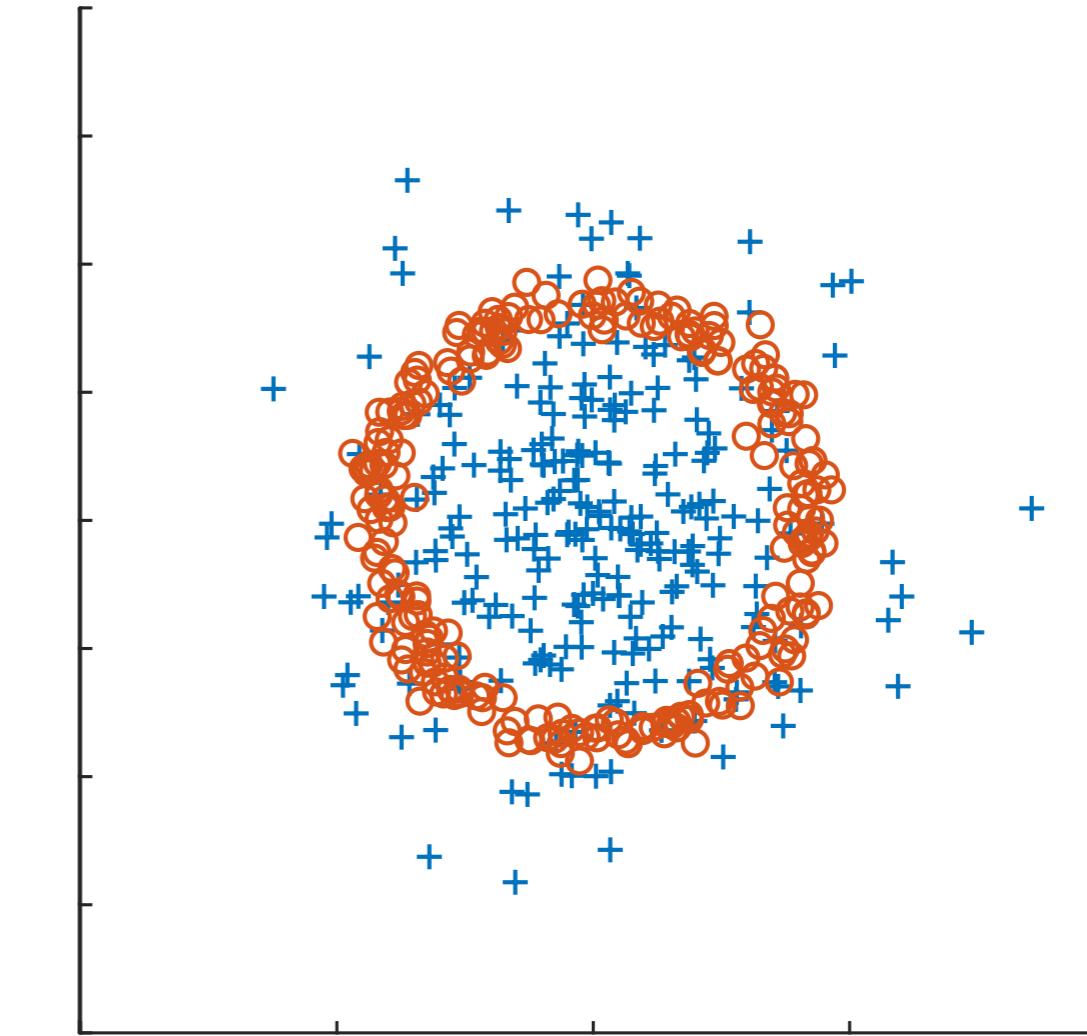
GP

What is
time series analysis?

unsupervised

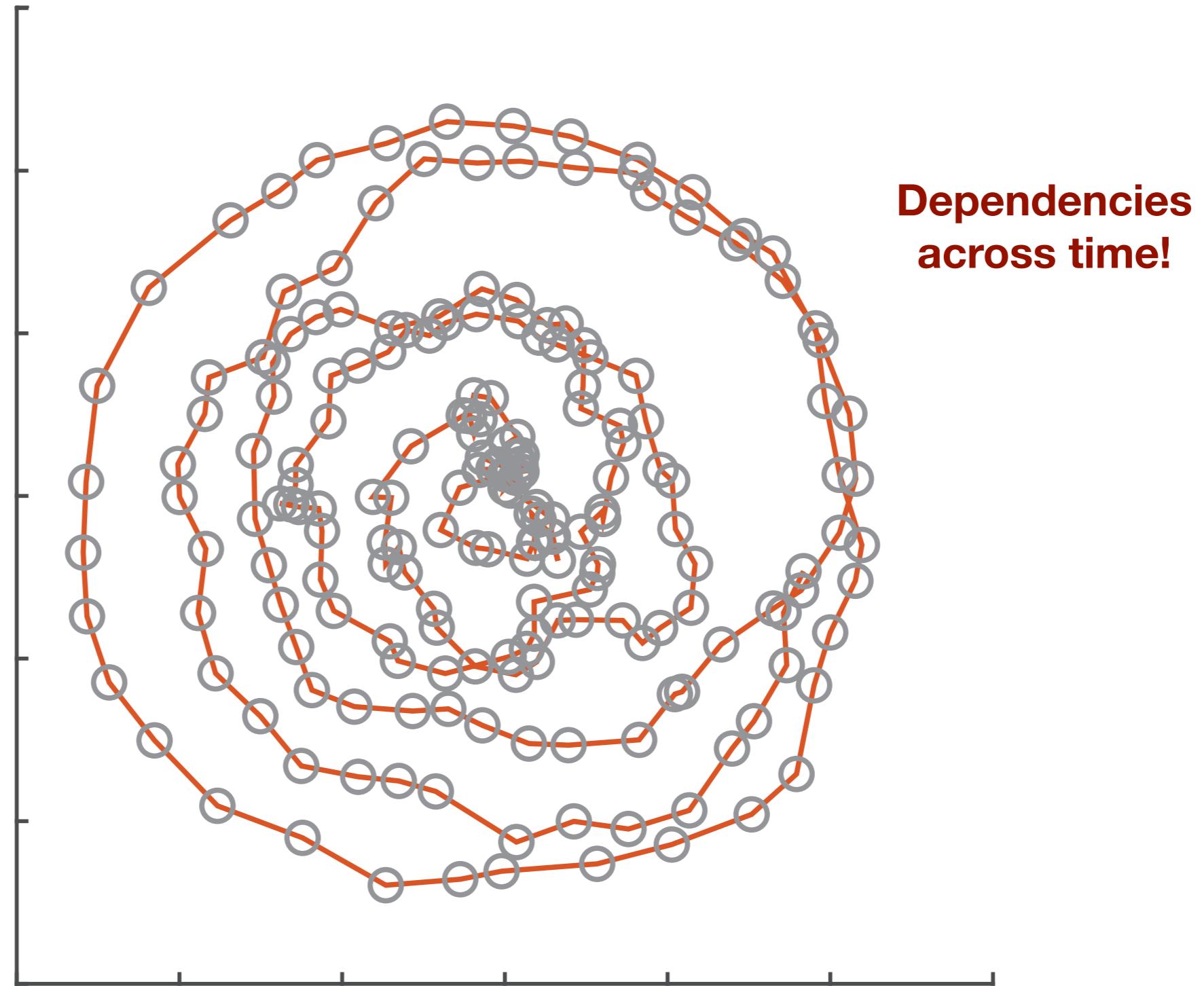


supervised



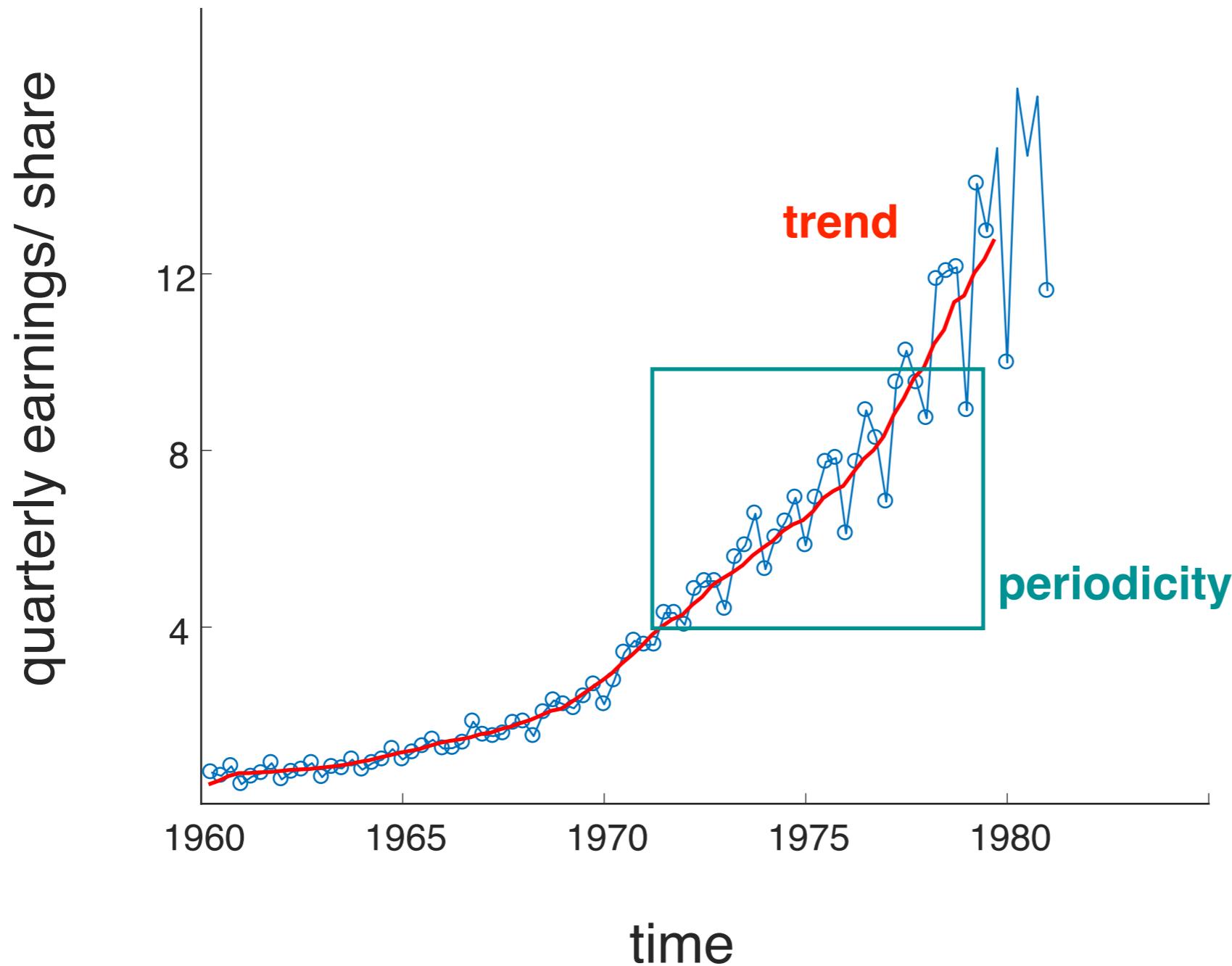
All ML: finding structure and using it to make predictions

TSA: finding structure and using it to make predictions in sequential data



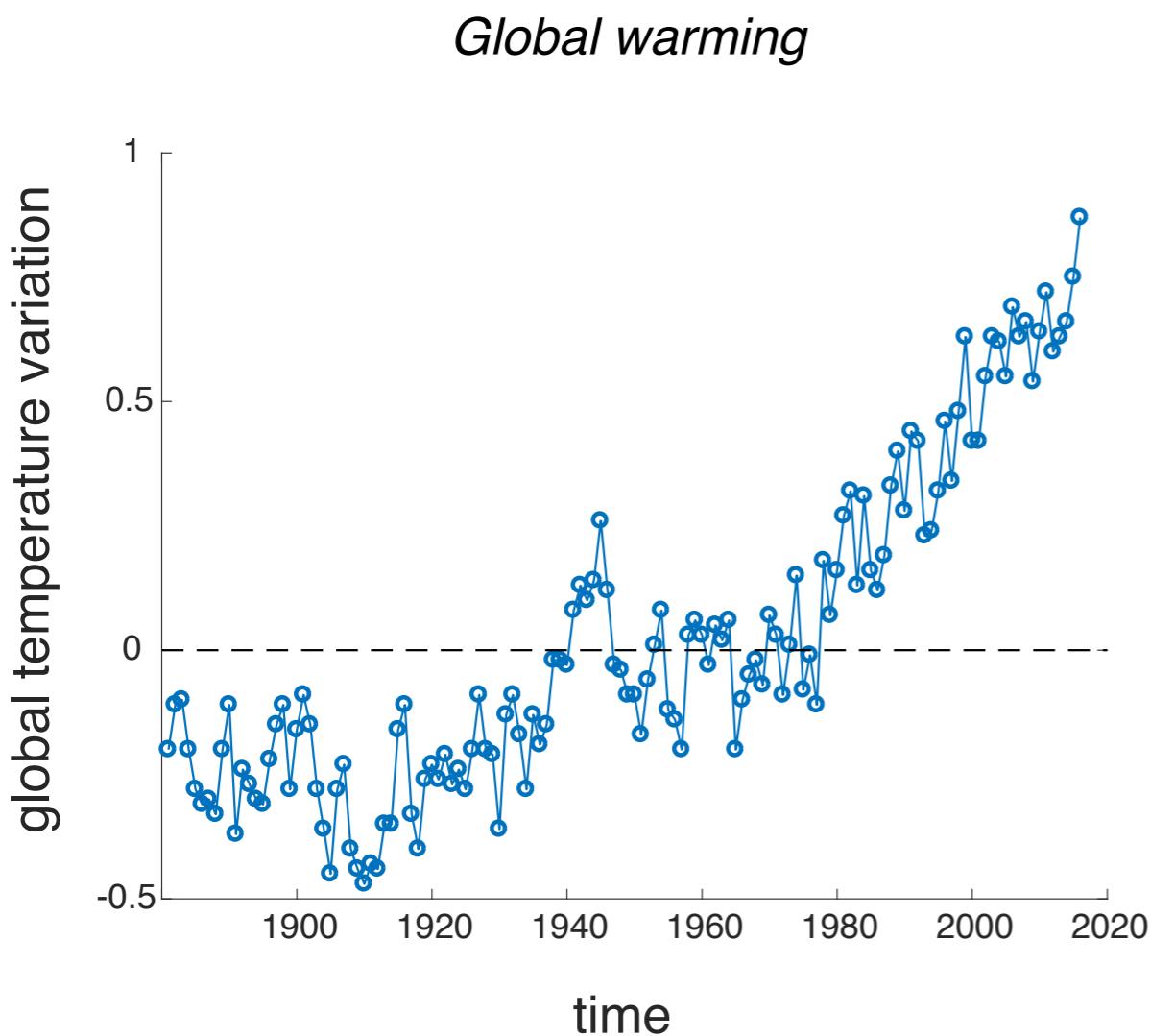
Some concrete examples

Task: predict future earnings, interpret data structure

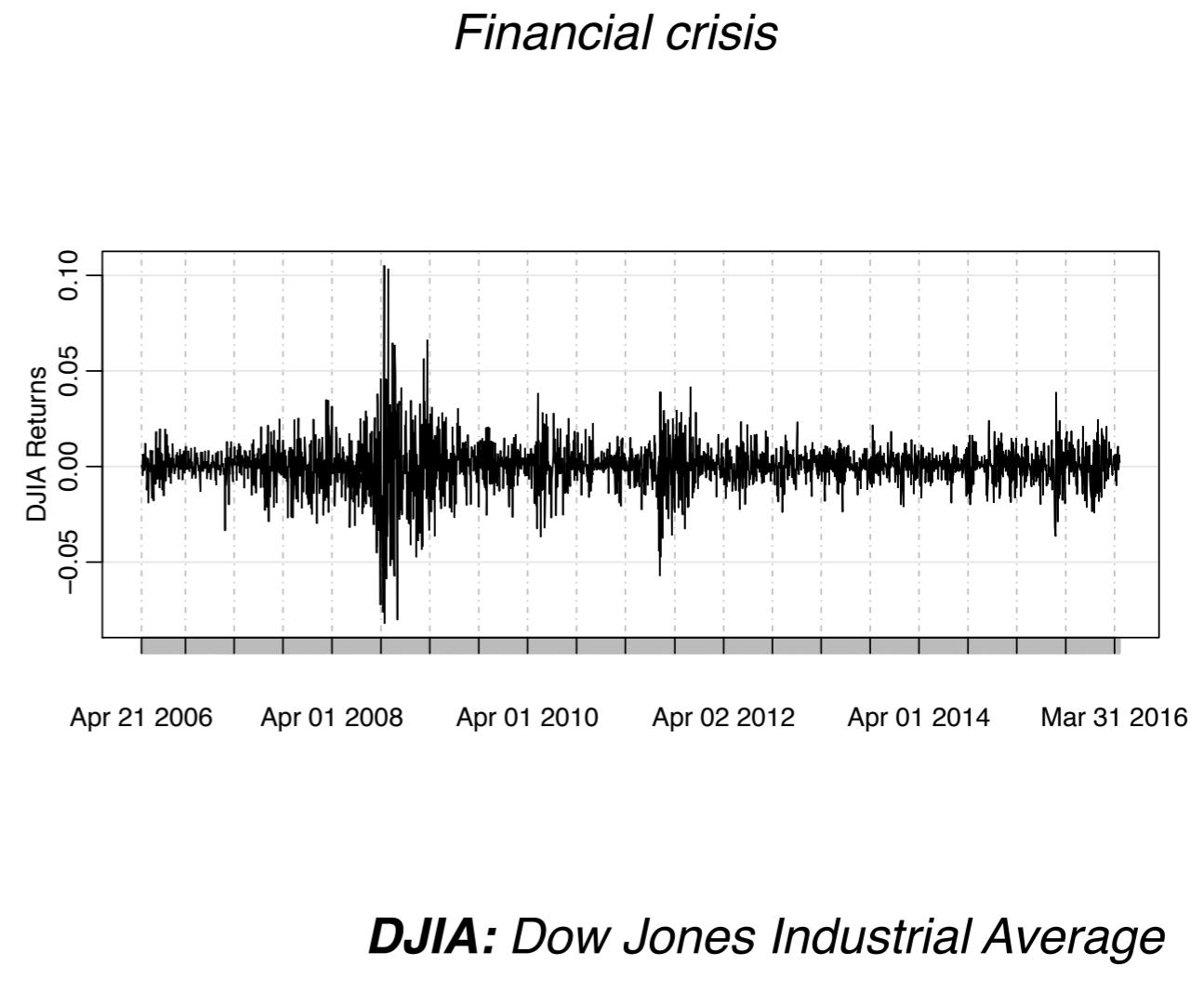


Johnson & Johnson

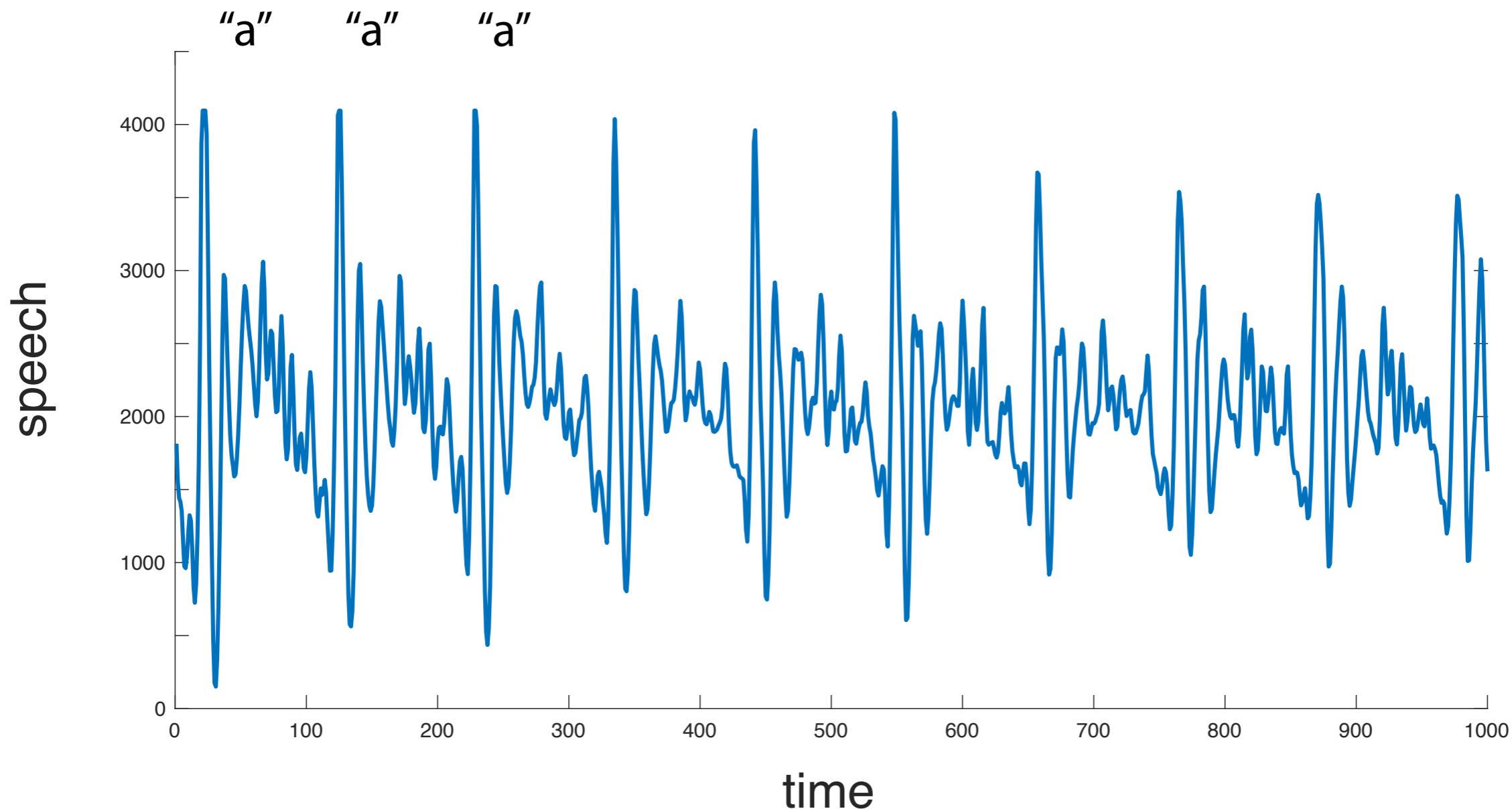
**Task: identifying trends,
do things change
systematically over time?**



Task: detect high volatility periods

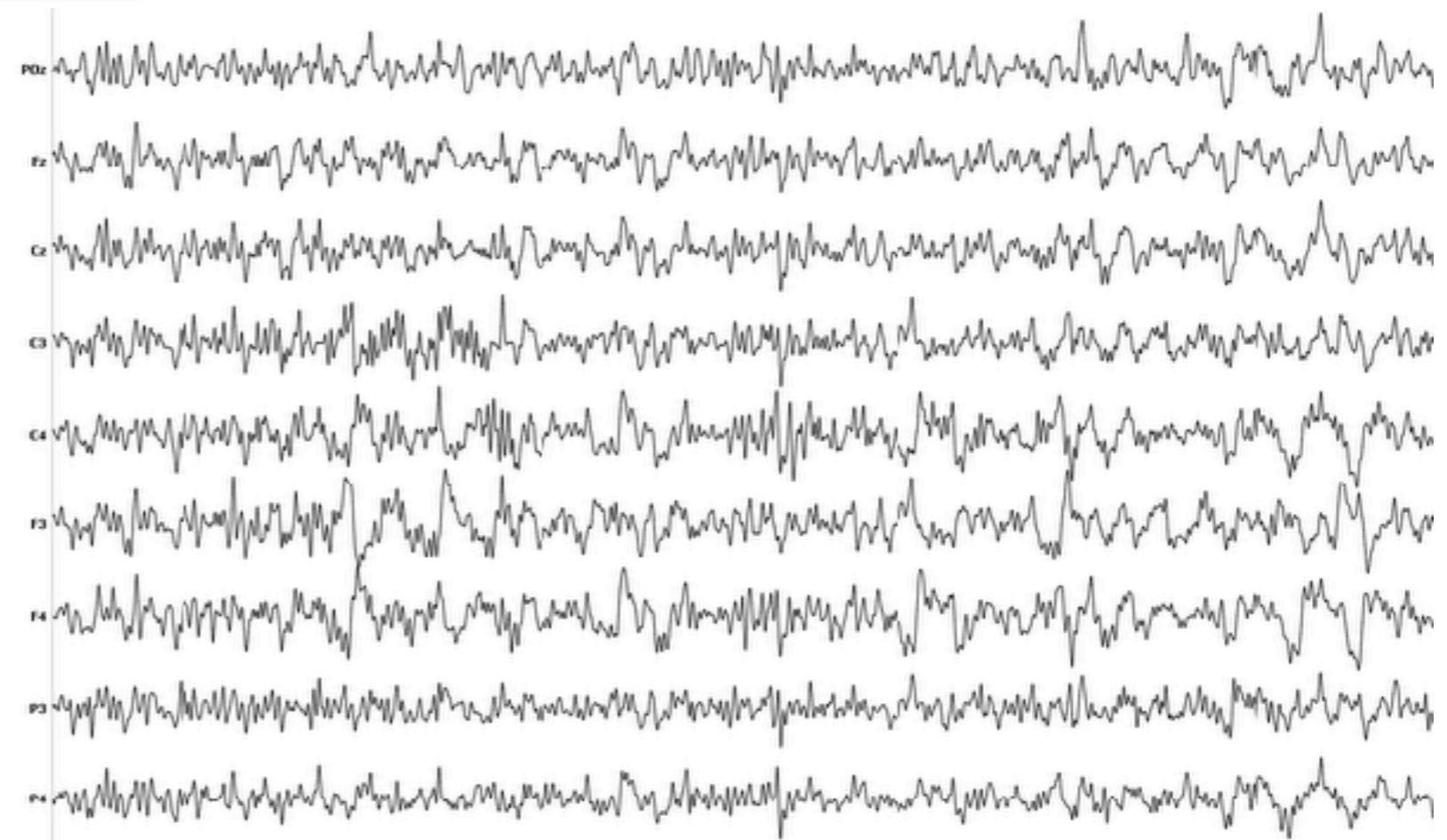


Task: infer discrete latent structure from analog signals





Multivariate time series: EEG

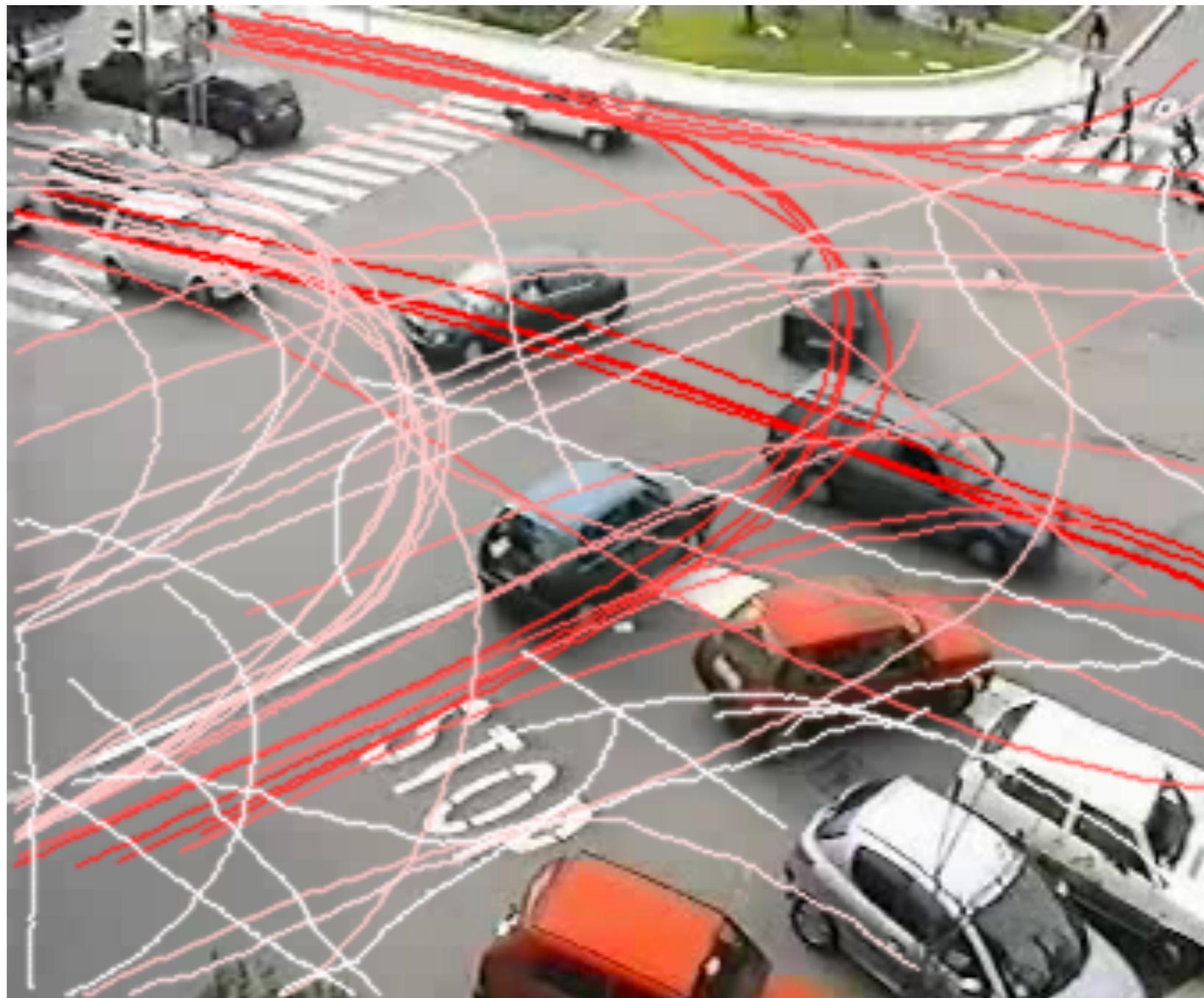




Task: decoding Brain-computer interfaces



Self driving cars



Tasks: simulation, control

We need uncertainty representation for
optimal decision making, risk minimization

Intro: what is a time series?

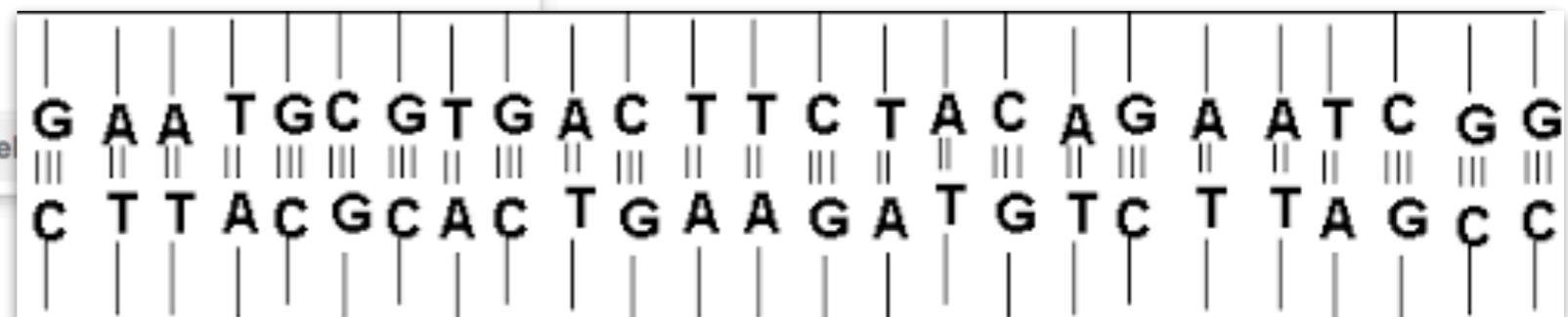


time se|

- time series analysis
- time series
- time sensitive
- time seconds
- time server
- time sert
- time series database
- time served
- time series graph
- time series regression

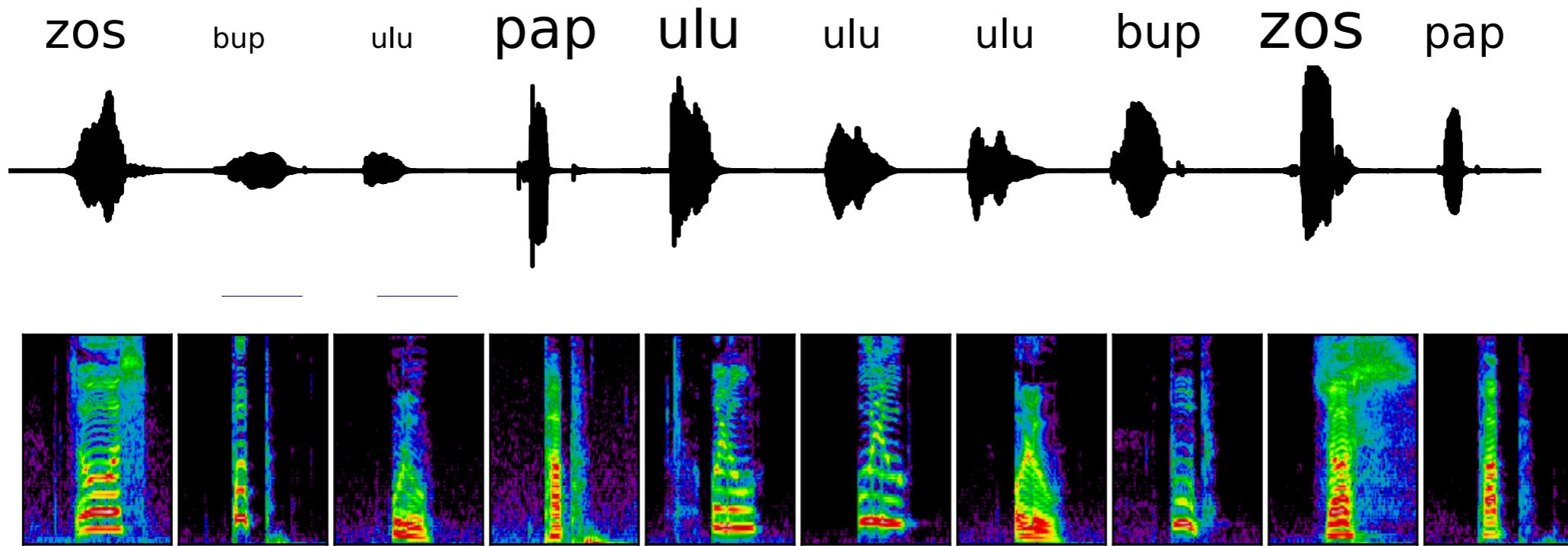
Google Search I'm Feeling Lucky

Language!
NLP
Machine translation



More general sequential structure:
e.g. sequence of nucleotide base pairs in DNA

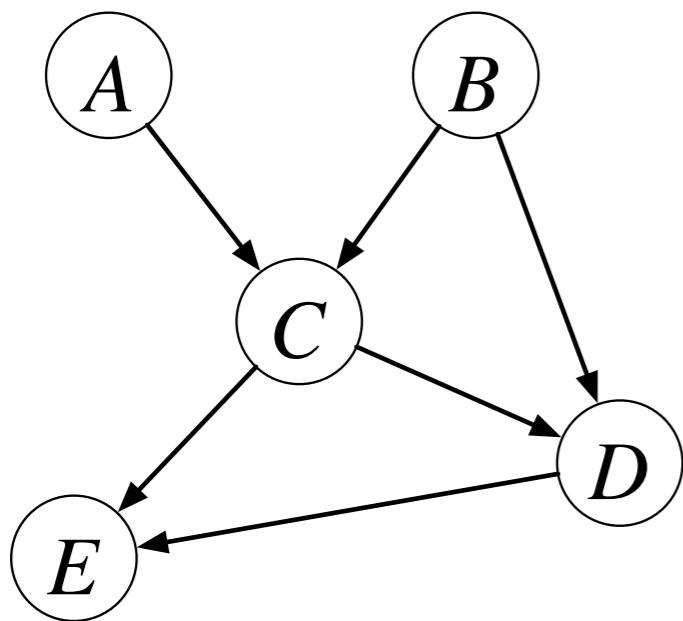
Intro: what is a time series?



Model data in the frequency domain
e.g. automated speech recognition

Tasks: identify latent structure, denoising...

Probabilistic time series analysis



1. Define generative models capturing relevant statistical structure in the data
(vs. recognition models)

Depiction: **graphical model**

$$p(A, B, C, D, E) = p(A)p(B)p(C|A, B)p(D|B, C)p(E|C, D)$$

2. Fit models to data

3. Make predictions

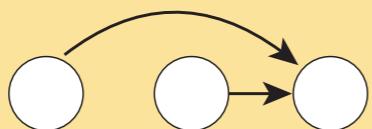
Overview of the course

TIME

Lagged relationships

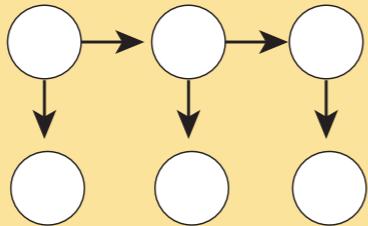
**Model temporal
dependencies directly**

ARIMA



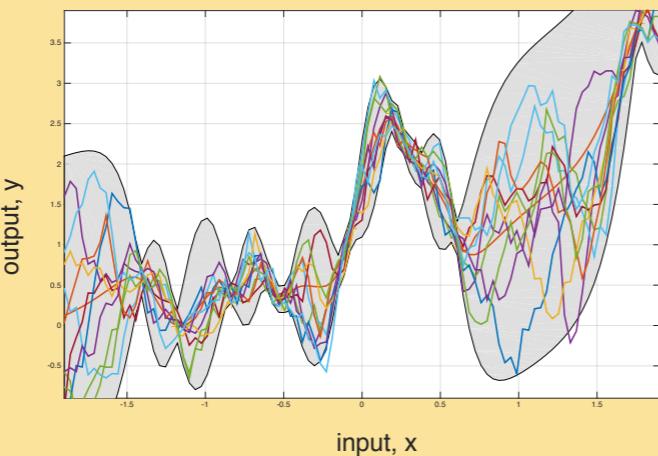
Latent structure

LDSs, HMMs



Advanced topics

GPs, RNNs



FREQUENCY

Periodic structure

**Stationary processes
seasonality/periodicity**

**Probabilistic
spectral analysis**

Mixing it up

Non stationary
spectral structure (?)

**Guest lecture:
example state of the art
This year: Andrew Wilson**

General philosophy of the course:

- Strong foundations for the classic models
- Understand generalizations and links between models

In lecture 1 we will:

- Review example applications**
- Basic statistics for time series**

Intro: what is a time series?

Definitions. Basic statistics

What is a time series?

Formally, a collection of random variables indexed by time, t*

**Usually discrete time (digital data collection), but continuous time can be convenient in some cases*

$$\{X_1, X_2, \dots, X_t \dots\}$$

“stochastic process”
data = “realisation”

Unlike the traditional case, NOT I.I.D. !!!

These **dependencies** are the main point; it's what makes prediction possible.

Fully specified by joint*:

$$P(X_1 \leq x_1, \dots, X_t \leq x_t \dots)$$

These **dependencies** are also the main problem: they make the math hard.

***Intractable in general, we limit ourselves to more structured classes of distributions*

Basic statistics of a time series

Mean

$$\mu_X(t) = \mathbb{E}(X_t)$$

Covariance

$$R_X(t, u) = \text{cov}(X_t, X_u)$$

**Auto-Correlation Function
(ACF)**

$$\rho_X(t, u) = \frac{R_X(t, u)}{\sqrt{R_X(t, t), R_X(u, u)}}$$

measures linear predictability of X_t from X_s

$$-1 \leq \rho_X(t, u) \leq 1$$

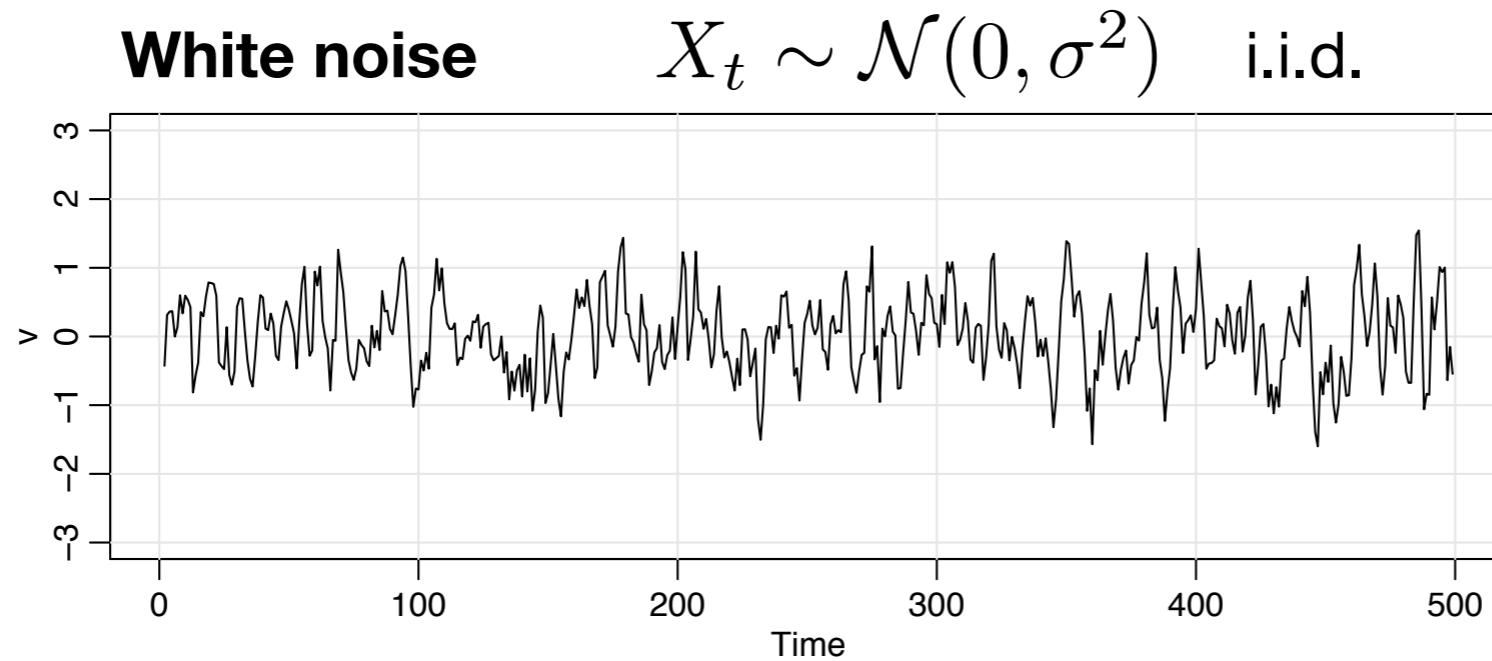
Cross-Covariance

$$R_{X,Y}(t, u) = \text{cov}(X_t, Y_u)$$

**Cross-Correlation Function
(ACF)**

$$\rho_{X,Y}(t, u) = \frac{R_{X,Y}(t, u)}{\sqrt{R_{X,Y}(t, t), R_{X,Y}(u, u)}}$$

Example stochastic processes



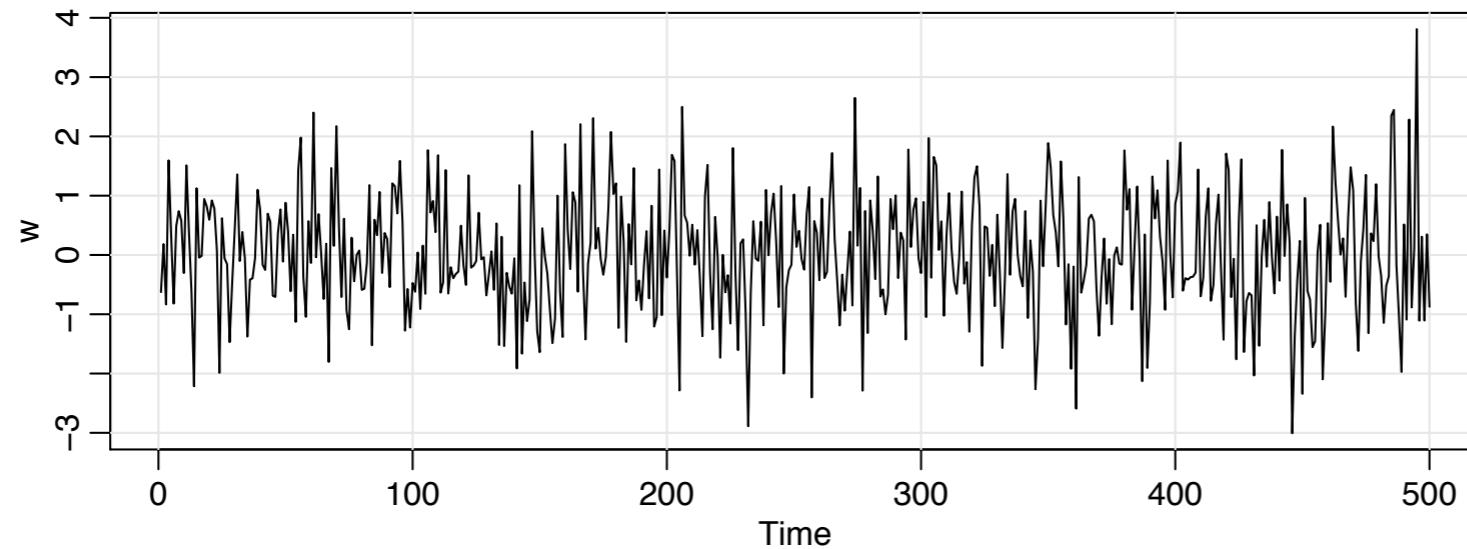
Trivially, white noise has

$$\mu_X(t) = 0$$

$$R_X(t, u) = \begin{cases} \sigma^2, & t = u \\ 0, & t \neq u \end{cases}$$

Example stochastic processes

Moving average (MA) $v_t = \frac{1}{3} (w_{t-1} + w_t + w_{t+1})$



filtered white noise

$$\mu_V(t) = 0$$

$$R_V(t, u) = \begin{cases} 1/3 \sigma^2 & , t = u \\ 2/9 \sigma^2 & , |t - u| = 1 \\ 1/9 \sigma^2 & , |t - u| = 2 \\ 0, & |t - u| > 2 \end{cases}$$

***Useful: Cov. of linear combinations**

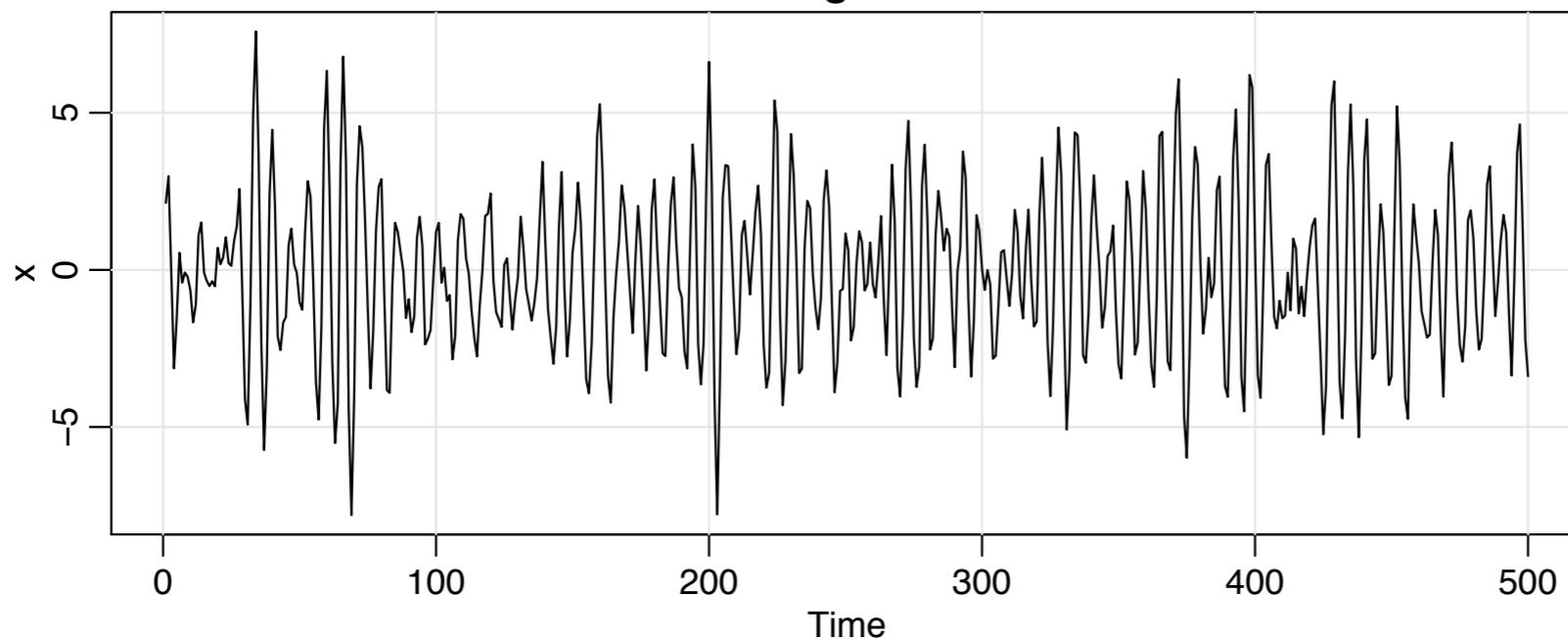
$$U = \sum_i a_i X_i$$

$$V = \sum_i b_i Y_i$$

$$\text{cov}(V, U) = \sum_{i,j} a_i b_j \text{cov}(X_i, Y_j)$$

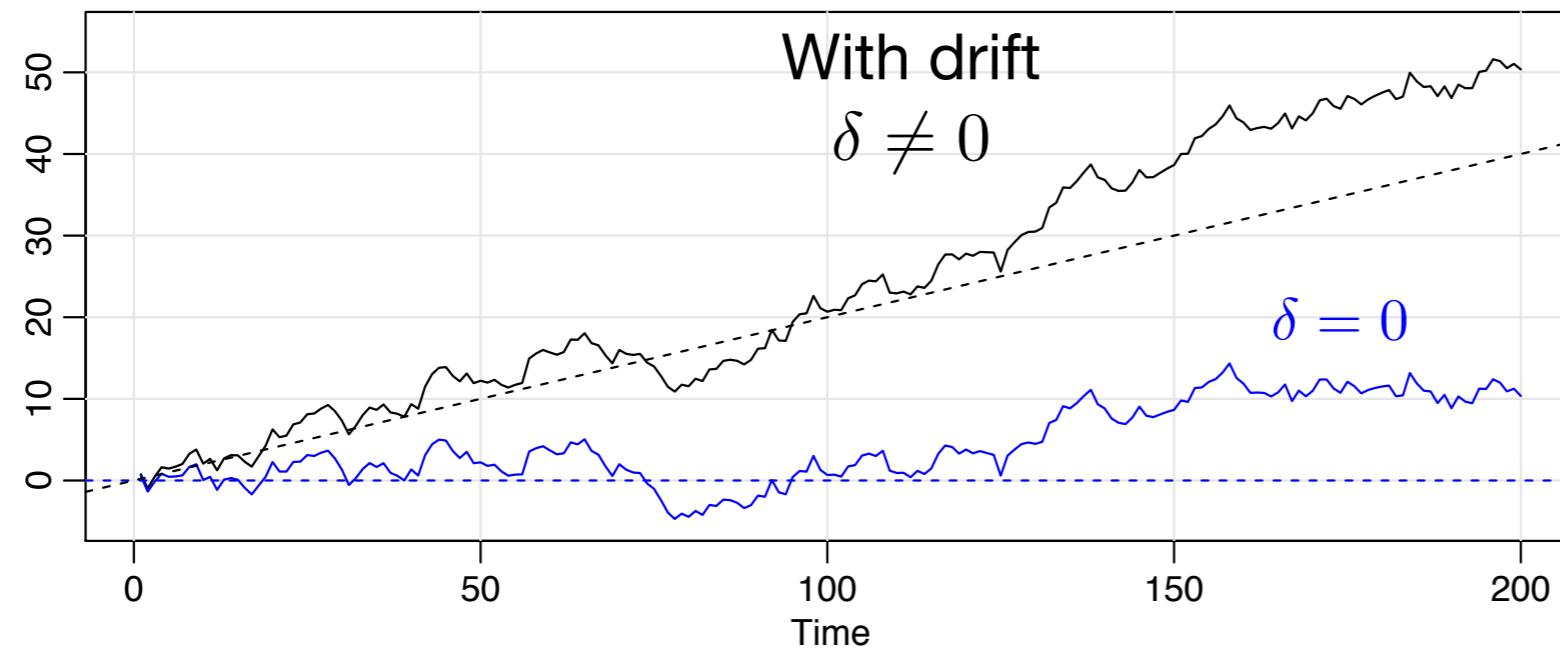
Autoregressive process (AR)

$$x_t = x_{t-1} - 0.9x_{t-2} + w_t$$



These are simple examples of ARIMA models – discussed in L2 .

Random walk $x_t = \delta + x_{t-1} + w_t$



If we unfold recursion:

$$x_t = t\delta + \sum_{i \leq t} w_i$$

$$\mu_X(t) = t\delta$$

$$R_X(t, u) = \min(t, u)\sigma^2$$

Cross-Covariance

$$R_{X,Y}(t, u) = \text{cov} (X_t, Y_u)$$

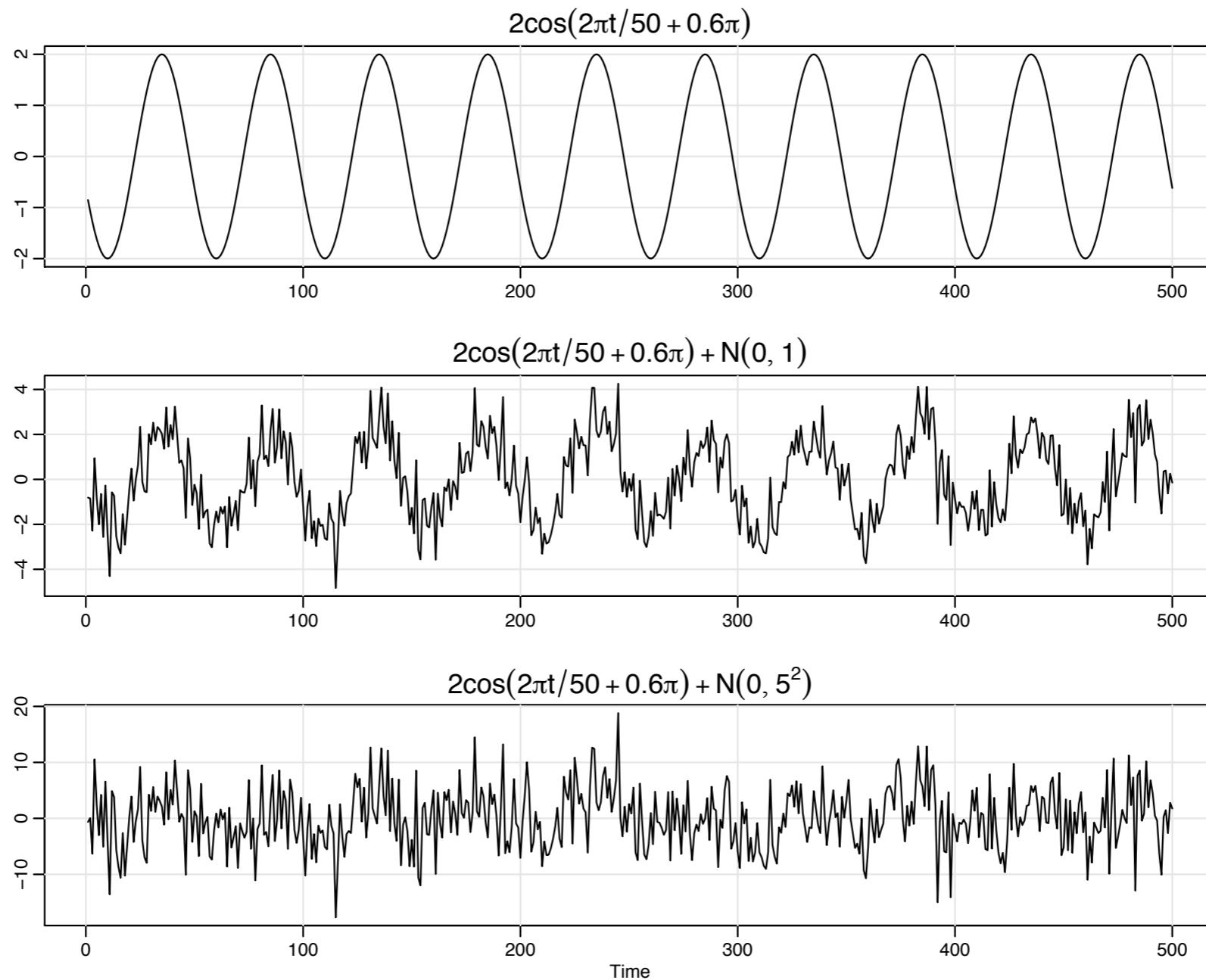
**Cross-Correlation Function
(ACF)**

$$\rho_{X,Y}(t, u) = \frac{R_{X,Y}(t, u)}{\sqrt{R_{X,Y}(t, t), R_{X,Y}(u, u)}}$$

E.g. $x_t = w_t + w_{t-1}$ and $y_t = w_t - w_{t-1}$,

$$\rho_{xy}(h) = \begin{cases} 0 & h = 0, \\ 1/2 & h = 1, \\ -1/2 & h = -1, \\ 0 & |h| \geq 2. \end{cases}$$

Signal vs noise



denoising in frequency domain,
inferring latent structure in temporal domain

Basic statistics of a time series

“Strong stationarity”

$$\begin{aligned} & \{X_t, \dots, X_{t+K}\} \\ & \{X_{t+h}, \dots, X_{t+h+K}\} \end{aligned}$$

Identically distributed subsets
for all t,h,K

Consequences:

For single variables (K=0) this implies same marginals everywhere

$$P(X_t < x) = P(X_{t+h} < x) \text{ for all } t, h, \text{ and so } \mu_X(t) = \text{cte}$$

For single variables (K=1) this implies same pairwise dependencies

Basic statistics of a time series

“(Weak) stationarity”

$$\mu_X(t) = \text{const.}$$

$$R_X(\Delta t) = \text{cov}(X_t, X_{t+\Delta t})$$

+finite variance

Example: moving averages

A strongly stationary process with finite variance is weakly stationary

Converse is more complicated:
a gaussian weakly stationary process is strictly stationary

*Note: change in notation, for stationary processes R_x has a single argument

Basic statistics of a time series

“Trend stationarity”

$$\mu_X(t) \neq \text{const.}$$

$$R_X(\Delta t) = \text{cov}(X_t, X_{t+\Delta t})$$

This means that data can be partitioned into a time-dependent term + zero-mean stationary process

e.g. sigmoid + white noise

Final note: random walks are non-stationary

Basic statistics of a time series

Linear process

A general version of filtered white noise

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Causality

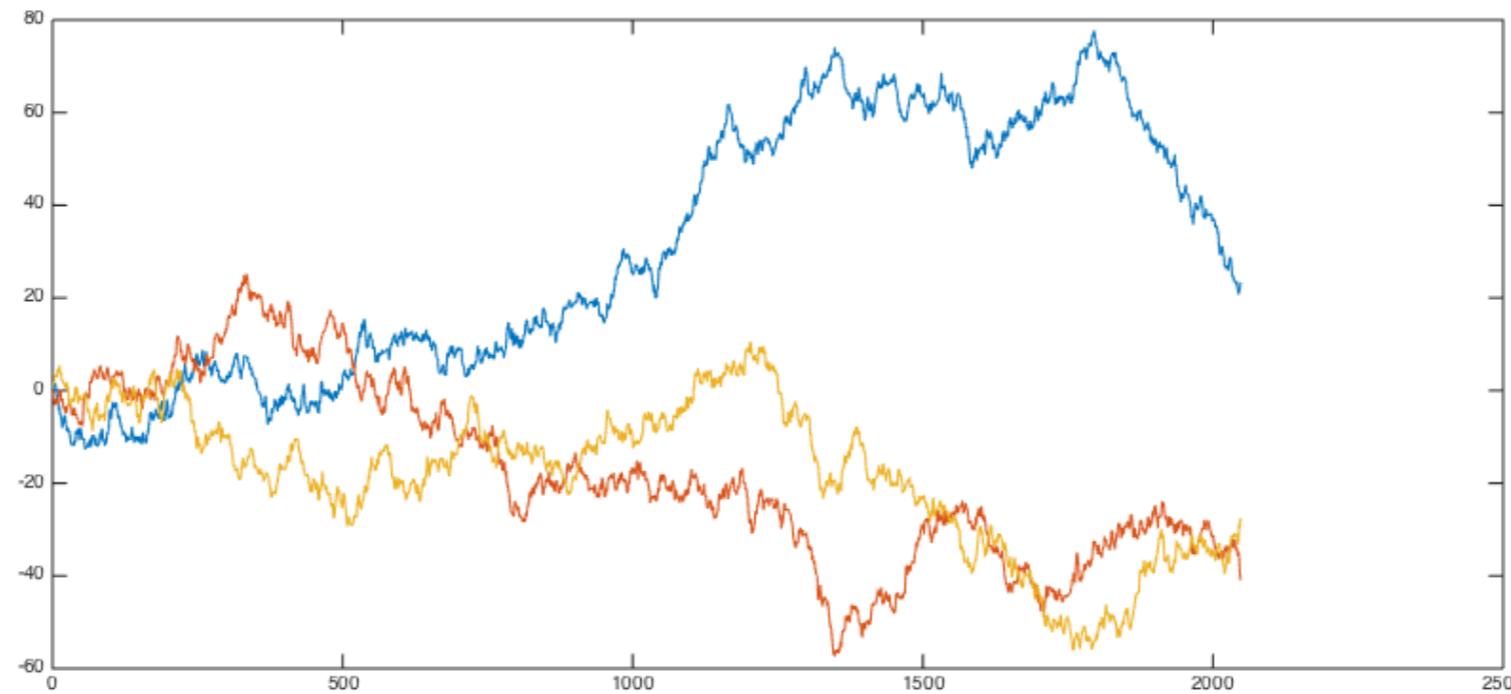
Present depends on past but not on future

It's often a natural assumption for real world data

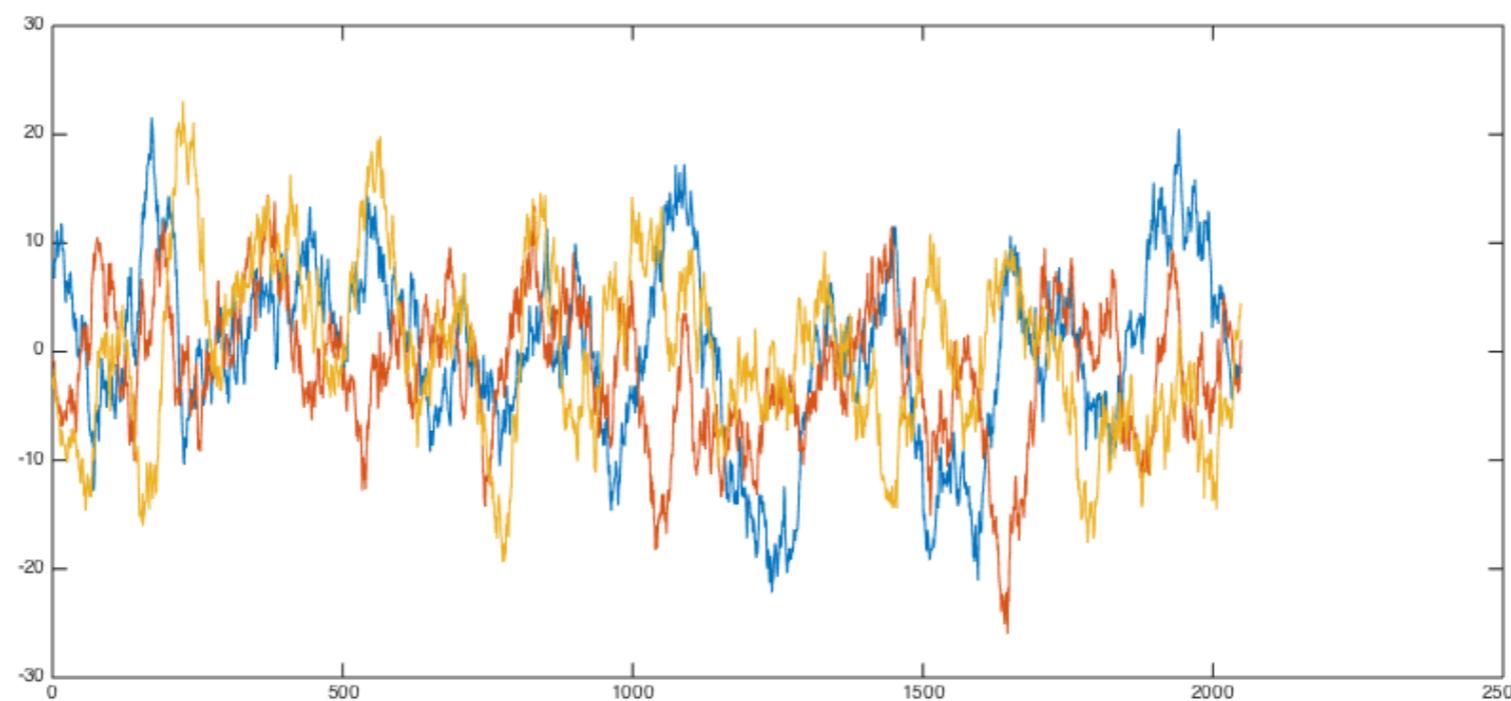
Not necessarily true for all sequential data (e.g. DNA)

Causal linear process:

$$\psi_j = 0 \text{ for future components } (j < 0)$$



Random walk
(non stationary)



stationary

Simpler structure,
Easier to estimate

Empirical measurements (**stationary** process)

$$\hat{\mu}_x = \frac{1}{T} \sum_t x_t$$

$$\hat{R}_x(\Delta t) = \text{cov}(x_t, x_{t+\Delta t})$$

NEXT WEEK: AR models