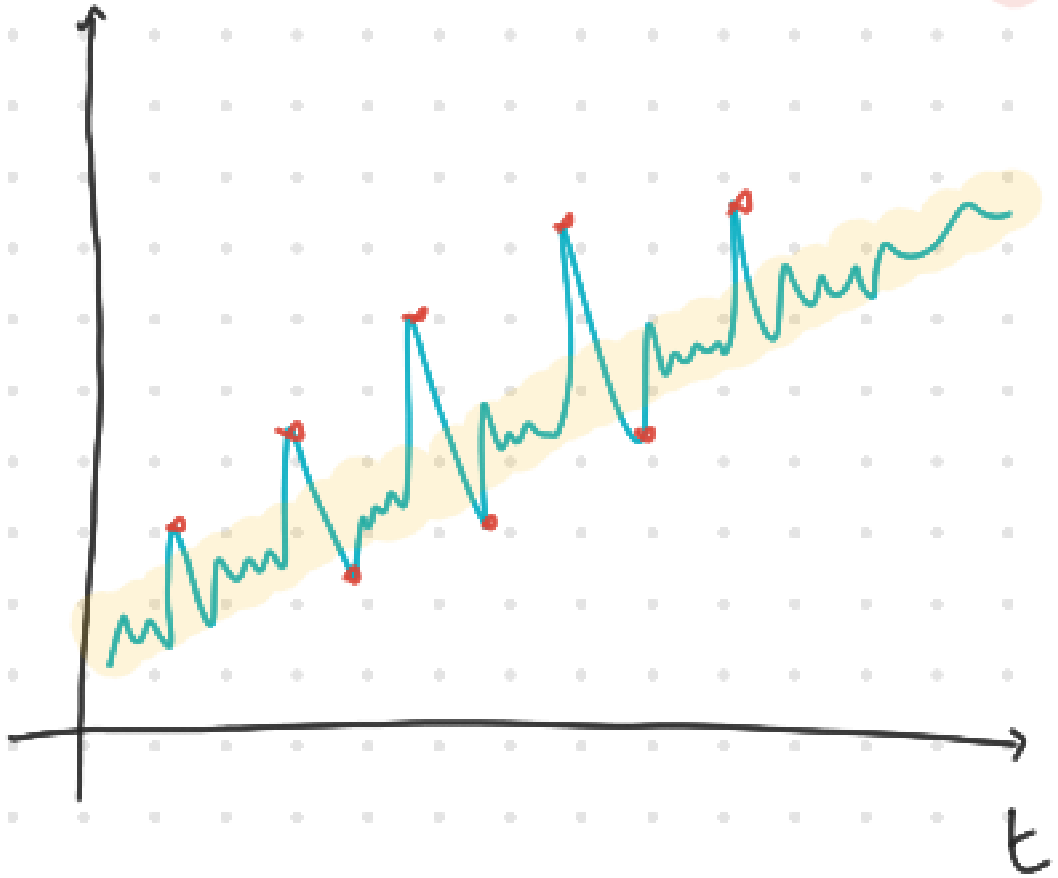
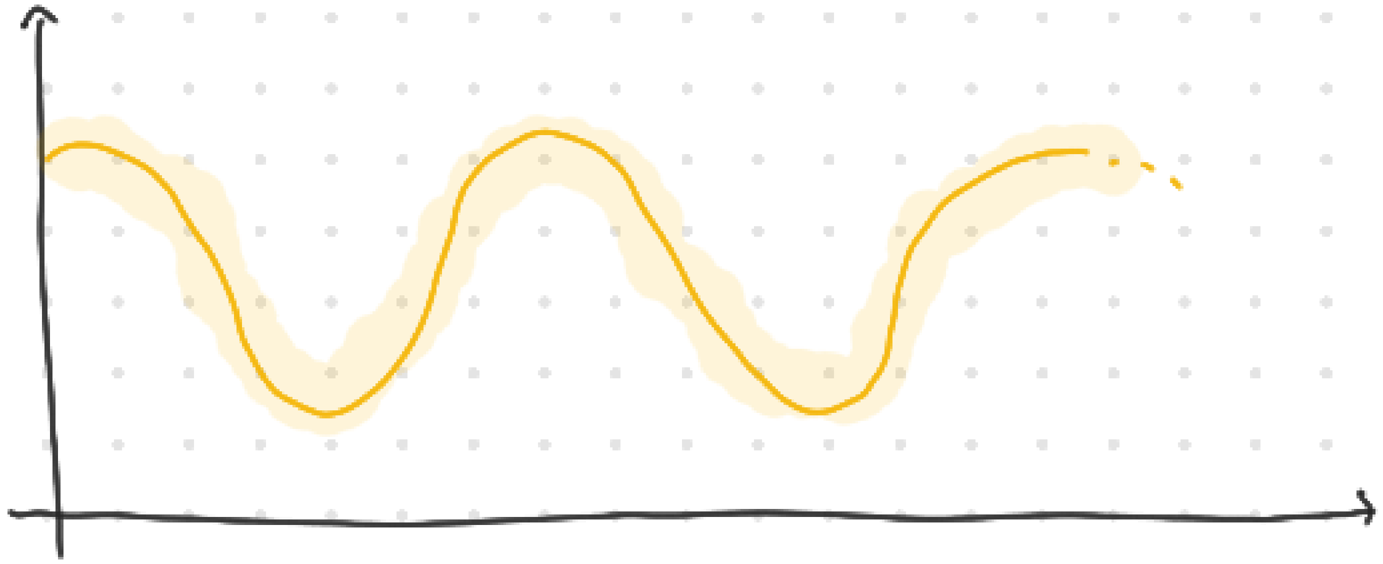
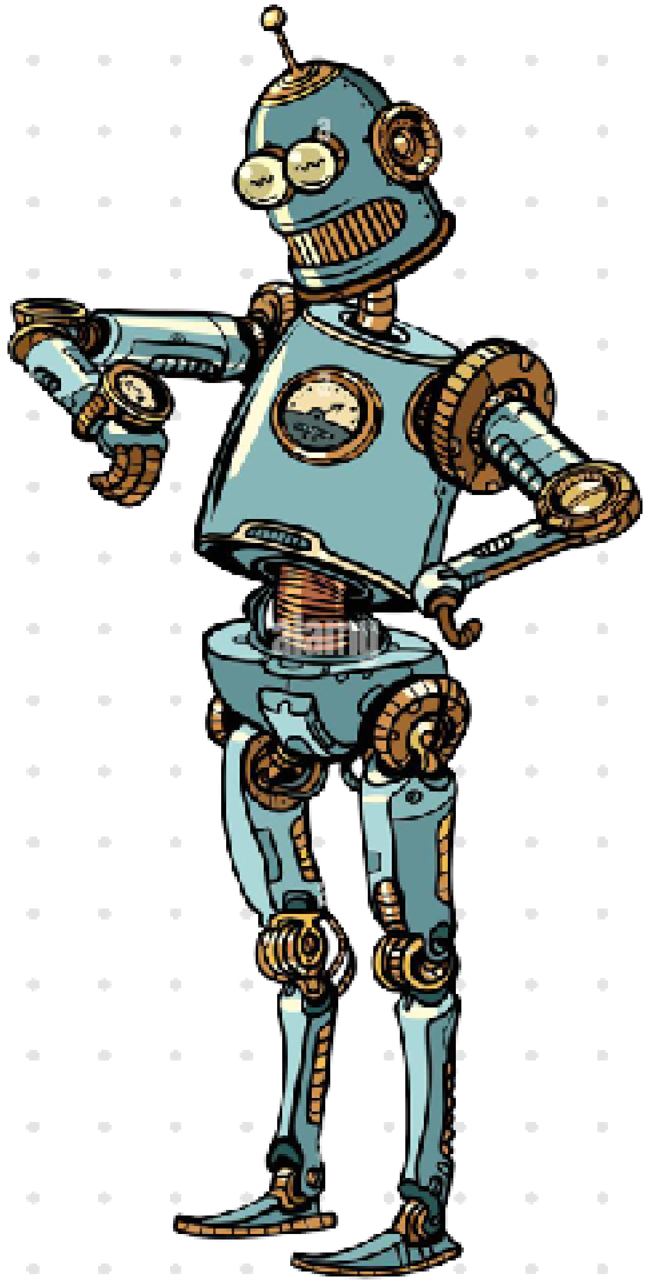
Time Series Analysis and Forecasting

by

Oscar de Felice







Experis
ManpowerGroup

Stochastic Process

The triplet $(\Sigma, \mathcal{F}, p)$ is called **PROBABILITY SPACE**

- Σ is a set called SAMPLE SPACE (non-empty)
- $\mathcal{F}$ is a σ-algebra over Σ
- p is a probability measure over $\mathcal{F}$.

$$\boxed{\begin{array}{l} \cdot\ \forall A \subseteq \Sigma,\ A \in \mathcal{F} \Rightarrow C(A) \in \mathcal{F} \\ \cdot\ \Sigma \in \mathcal{F} \\ \cdot\ \forall \{A_i\}_{i \in \mathbb{N}} \in \mathcal{F} \Rightarrow \bigcup_i A_i \in \mathcal{F} \end{array}}$$

A Random variable over $(\Sigma, \mathcal{F}, p)$ is a function

$$X : \Sigma \longrightarrow \mathbb{R} \ :\ \forall a \in \mathbb{R},\quad X^{-1}(]-\infty, a]) \in \mathcal{F}$$

measurable function

STOCHASTIC PROCESS

A stochastic process is a family of real random variables over the same probability space $(\Omega, \mathcal{F}, P)$

$$X := \{ X_i(\omega),\ i \in T \},$$

T is the index set

$T \subseteq \mathbb{Z}$ (discrete)

$T \subseteq \mathbb{R}$ (continuous)

Recall that each value a random variable assumes in $\mathbb{R}$, i.e. each $x(\omega)$, $\omega \in \Omega$ is called REALISATION of x.

A time series is a Realisation of a stochastic process

FINITE DIMENSIONAL DISTRIBUTION

Given $T := \{t_1, \dots, t_n\}$ Finite set of integers, $t_i \in \mathbb{Z} \quad \forall i$

we define a joint distribution function for $\underline{X} := \{\underline{X}_i(\omega) \mid i \in T\}$

$$\boxed{F_{t_1,\dots t_n}(x_{t_1}, \dots, x_{t_n}) := p\left(\underline{X}_{t_1}(\omega) \leq x_{t_1}, \dots, \underline{X}_{t_n}(\omega) \leq x_{t_n}\right)}$$

often denoted as $F_{\underline{X}}(x_{t_1}, \dots, x_{t_n})$

The finite dimensional distribution for stochastic process $\underline{X}$ is the set

$$\left\{ F_{\underline{X}}(x_{t_1}, \dots, x_{t_n}) \mid n \in \mathbb{N}, T \subset \mathbb{Z} \right\}$$

STATIONARITY

- A time sequence is stationary $\iff$ it's stationary the underlying stochastic process

- A stochastic process is (strong)-stationary $\iff$

$$F_X(x_{t_1+\tau}, \dots, x_{t_n+\tau}) = F_X(x_{t_1}, \dots, x_{t_n})$$

$$\forall n, \tau$$

▷ A stochastic process is (Weak)-stationary $\Longleftrightarrow$

1. First moment of x_i is constant $\mathbb{E}[x] = \mu$

2. The second moment of x_i is finite $\forall i$

3. The cross moment (covariance function) depends only on time-difference

$$\gamma_k := \operatorname{cov}(x_t, x_{t+k}) \quad \text{does not depend on } t.$$

This implies variance is always constant

$$\sigma^2_{x_t} = \operatorname{cov}(x_t, x_t) = \gamma_0$$

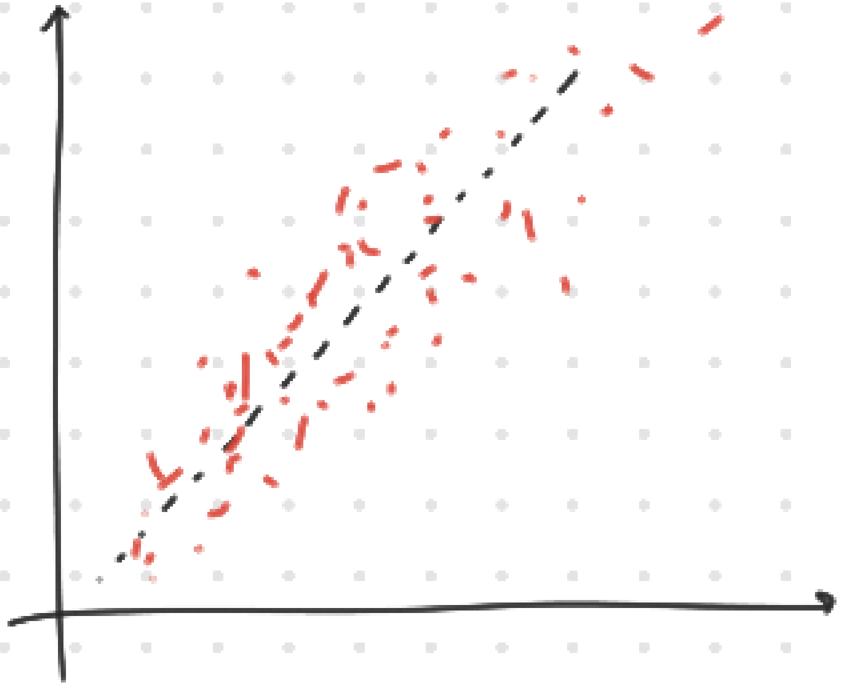
Experis
ManpowerGroup

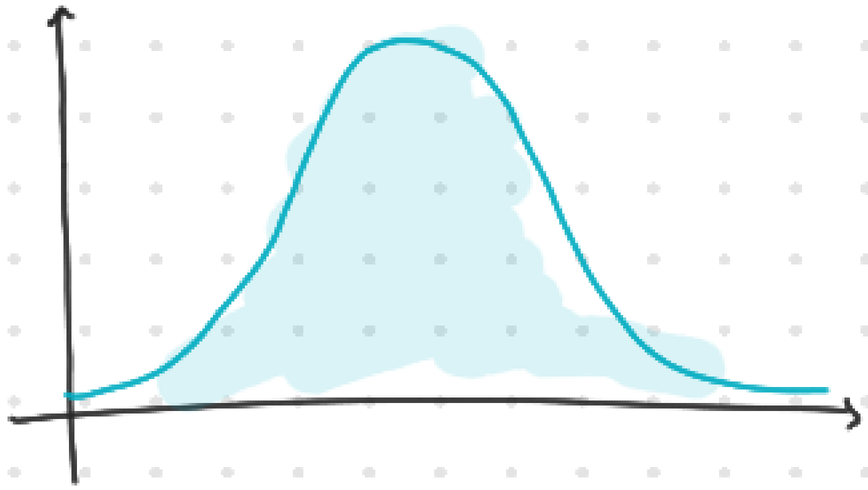
Some Statistics Review

1. Linear Regression
2. Normality of Residuals
3. Correlations

$$\boxed{y_i = (\beta_0 + \beta_1 x_i) + \varepsilon_i}$$

$$\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon)$$





$$\boxed{\rho(x, y) := \mathbb{E}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right)} = \frac{1}{\sigma_x \sigma_y} \operatorname{cov}(x, y)$$

STOCHASTIC PROCESSES

↳ Sequence of random variables $x_1, x_2, \dots, x_t$

each of them $\boxed{x_t \sim \Delta(\mu_t, \sigma_t)}$ distribution Δ

A time series is a realisation of a random process

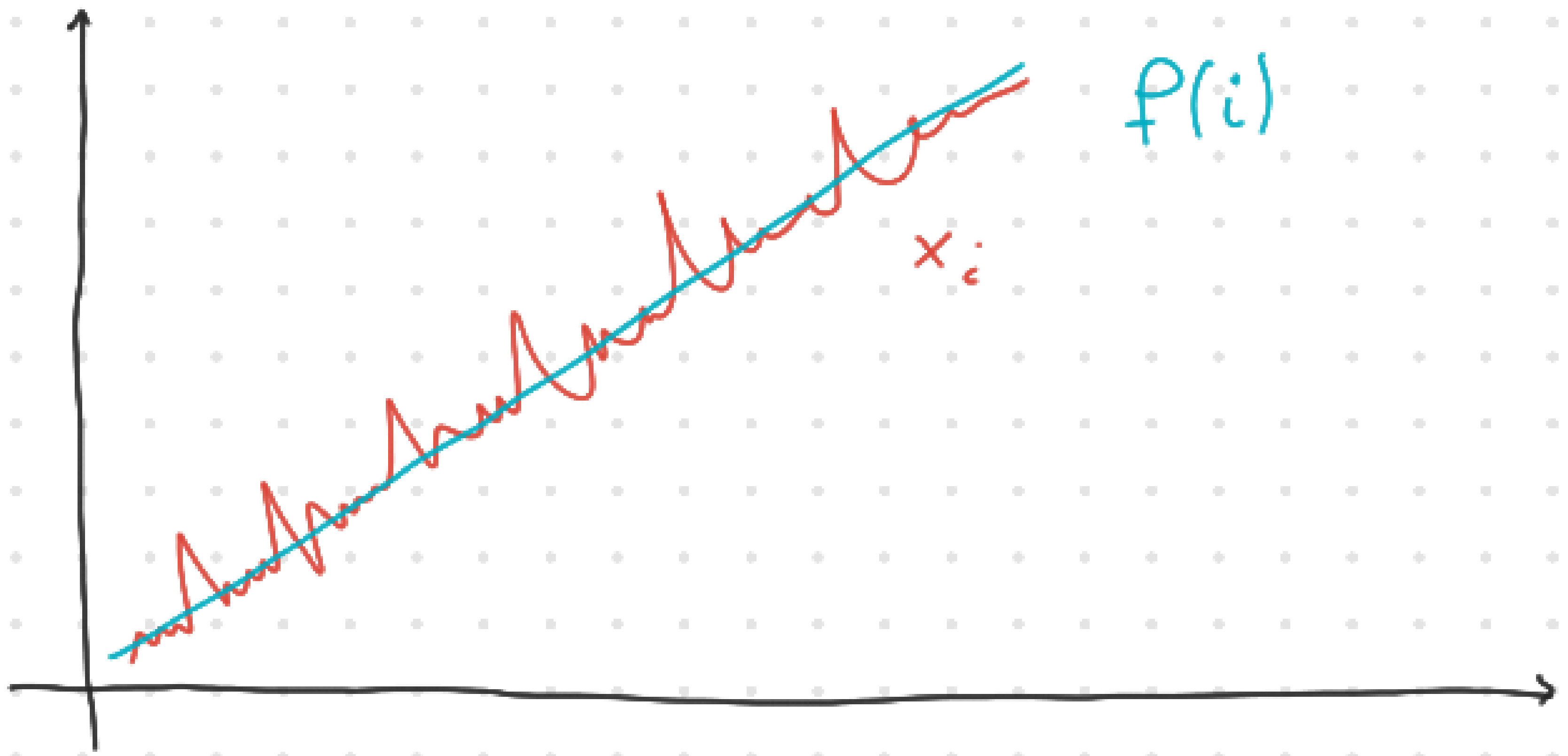
$$\begin{array}{ccccc} x_1, & x_2 \dots & & & x_t \\ \downarrow & \downarrow & & & \downarrow \\ \{ \hat{x}_1, & \hat{x}_2, & \dots & & \hat{x}_t, \dots \} \end{array}$$

▷ The main idea is to report any series to a stationary one → EASIER TO ANALYSE

Example: trend stationarity

Any process that can be expressed as

$$x_i = f(i) + \varepsilon_i \quad , \quad \varepsilon_i \text{ stationary series}: E[\varepsilon_i] = 0$$

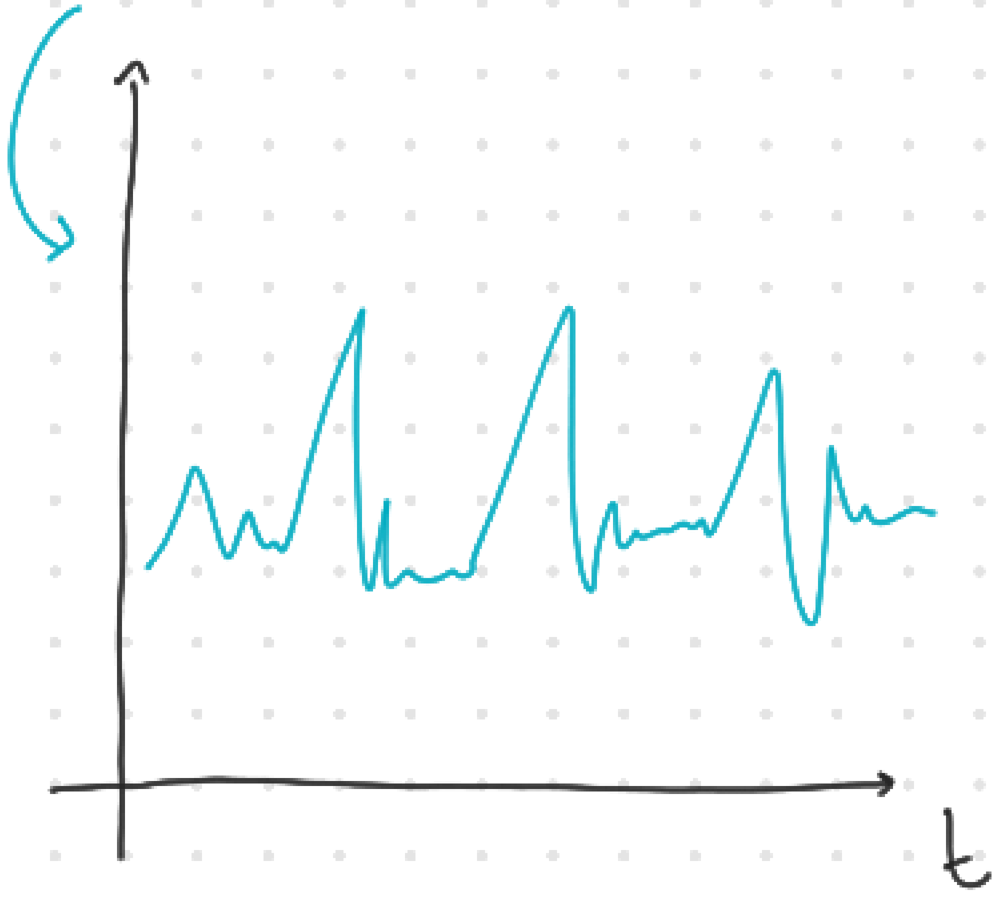


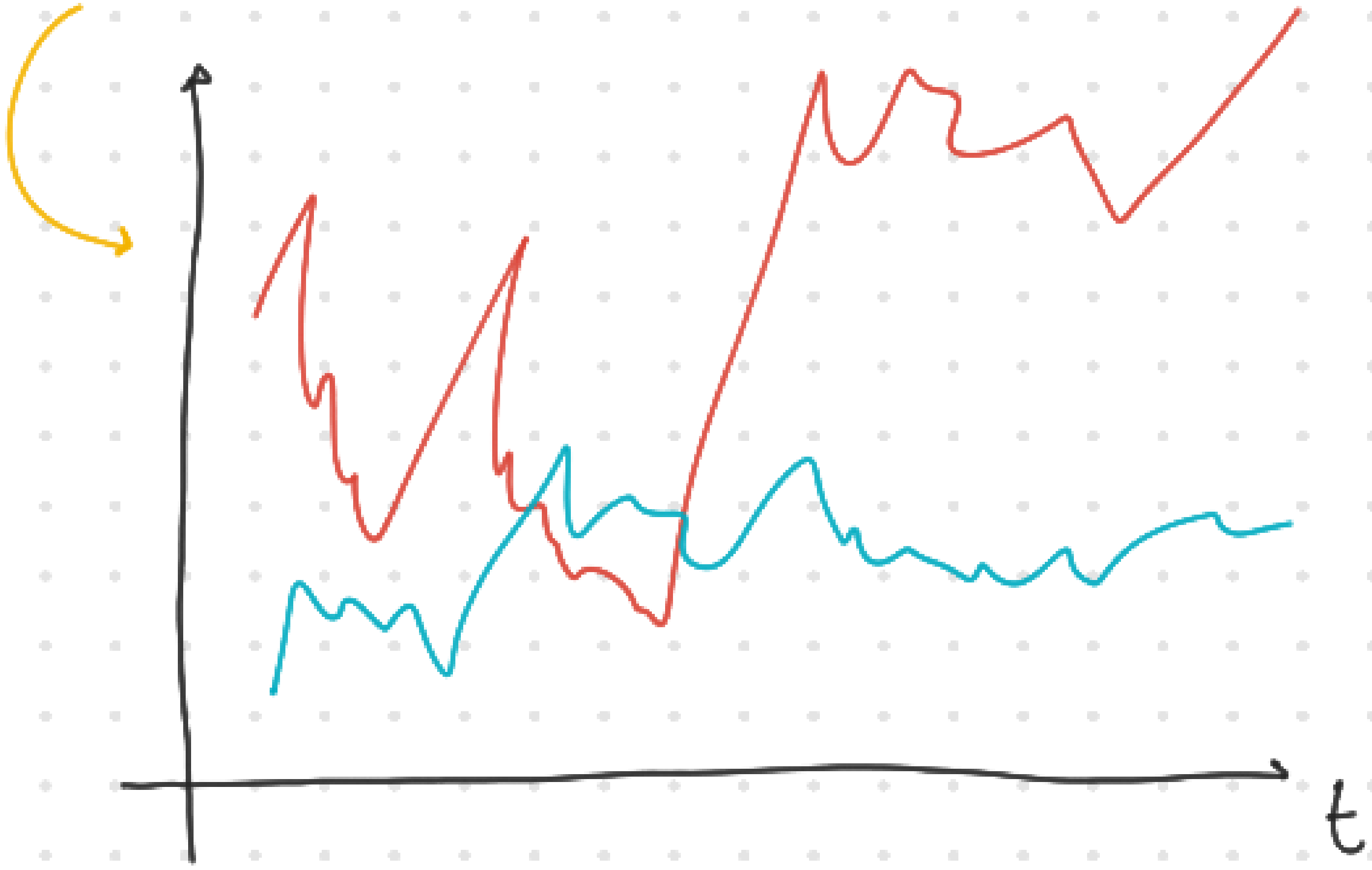
TIME SERIES ANALYSIS

Def. (Again)

"Sequence of data measured over time"

1. UNIVARIATE vs MULTIVARIATE





Examples:

→ Covid cases in one country

→ physical system pos
$x(t), y(t), z(t)$

TIME SERIES

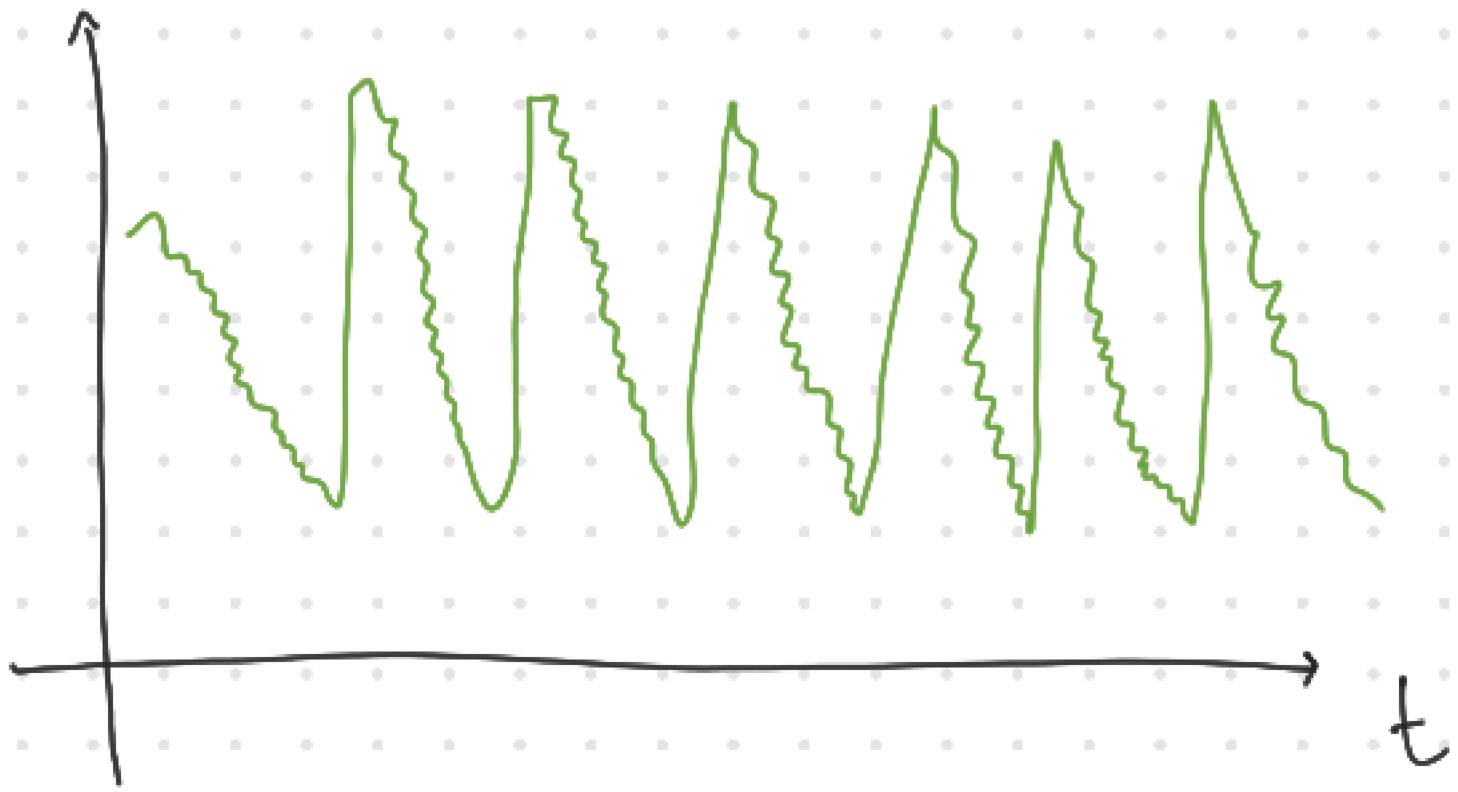
Definition: Series of data indexed in time order

SEQUENCE

$$\boxed{X_i \equiv X(t_i) \quad , \{t_i\}_{i=0}^{N}}$$

2. SEASONALITY

the tendency of a variable to behave in the same way periodically

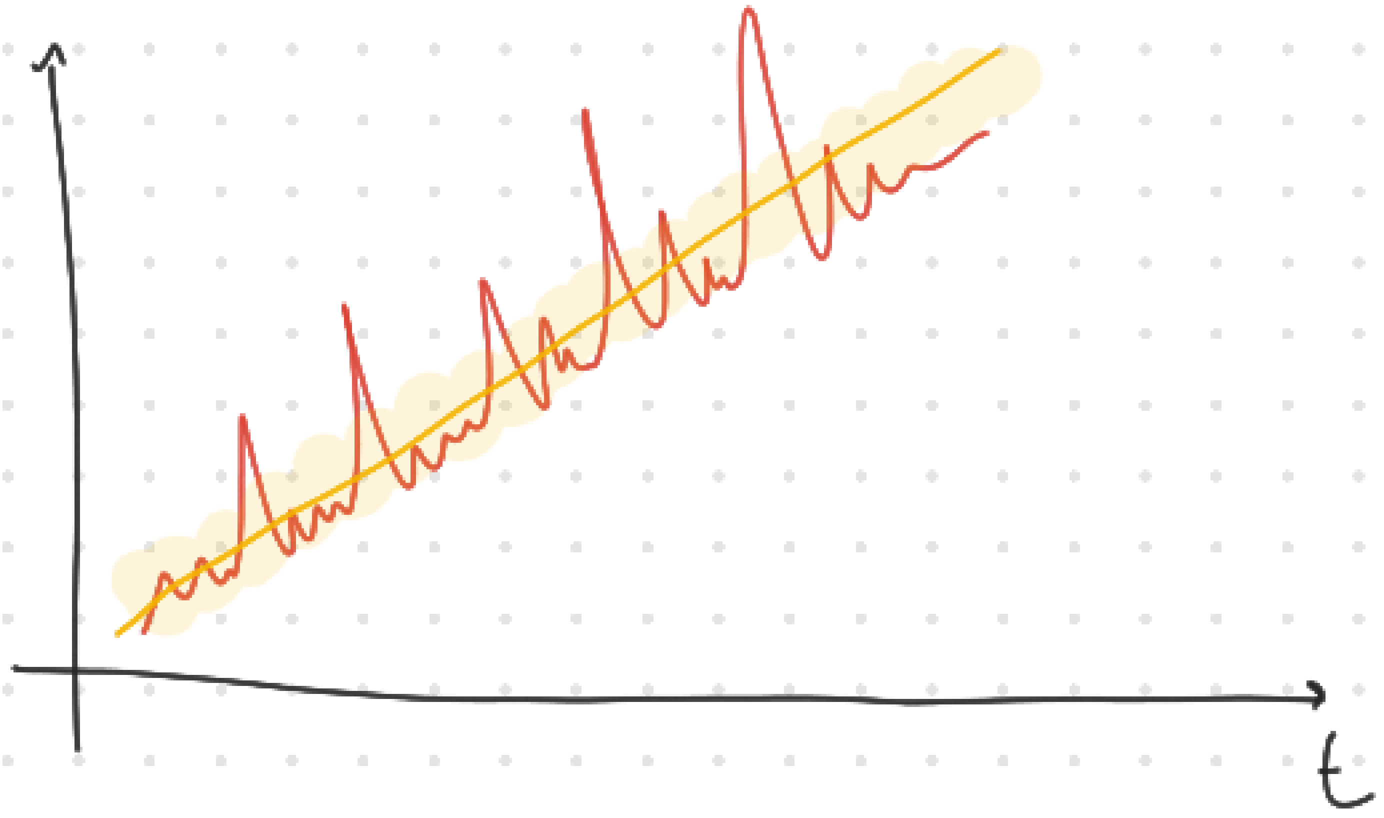


Example:

- Sales during christmast

3. TREND

↳ long term behaviour of a time-dep. variable



Example

- CO_2 concentration

Trend is something more stable than seasonality. ⟹ We need

DECOMPOSITION

TIME SERIES DECOMPOSITION

We have seen trend and seasonality as components of a time series. Meaning we can see it as a "sum" of different effect

Two models:

1. Additive model : $\boxed{x(t) = \sigma(t) + \tau(t) + \varepsilon(t)}$

2. Multiplicative model : $\boxed{x(t) = \sigma(t) \cdot \tau(t) \cdot \varepsilon(t)}$

FORECASTING

Some forecasting models:

- AR = AutoRegressive model → AR(p)

  Linear function of past values   memory based. lag p

$$\boxed{x(t) = \lambda + \phi_1 x(t-1) + \ldots + \phi_p x(t-p) + \varepsilon(t)}$$

- MA = Moving Average model ⟶ MA(q)

$$\boxed{x(t) = \varepsilon(t) + \vartheta_1 \varepsilon(t-1) + \ldots + \vartheta_q \varepsilon(t-q)}$$

moving average of the last (q+1) random shocks

- ARMA = AutoRegressive Moving Averege → ARMA(p,q)

$$\boxed{X(t) = \phi_1 x(t-1) + \ldots + \phi_p x(t-p) + \varepsilon(t) + \vartheta_1 \varepsilon(t-1) + \ldots + \vartheta_q \varepsilon(t-q)}$$

- ARIMA = AutoRegressive Integrated Moving Averege → ARIMA(d,p,q)

≡ ARMA(p+d, q)

$$x(t) = \phi_1 x(t-1) + \ldots + \phi_p x(t-p) + \varepsilon(t) + \vartheta_1 \varepsilon(t-1) + \ldots + \vartheta_q \varepsilon(t-q)$$

$$\Leftrightarrow \quad \boxed{\left(1 - \sum \phi_k L^k\right) x(t) = \left(1 + \sum_j \vartheta_j L^j\right) \varepsilon(t)}$$

where $L^k[x(t)] := x(t-k)$ LAG OPERATOR

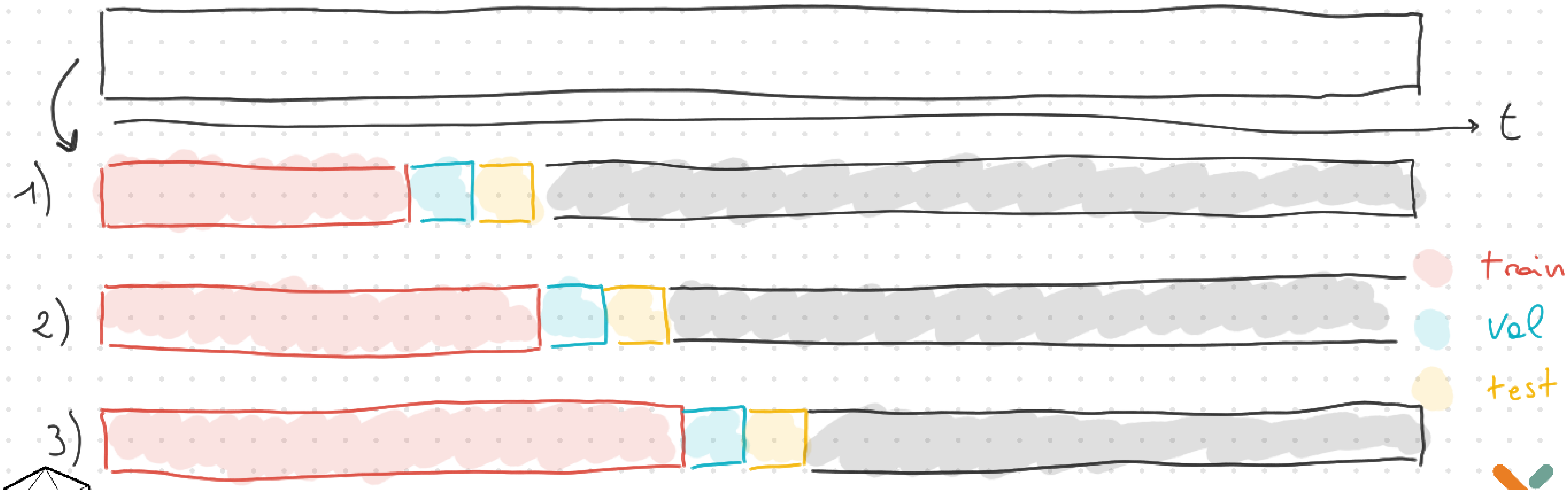
if the associated polinomial (coeff ϕ_k) has a root $s=1$ (mult. d)

we call the stochastic process INTEGRATED or DIFFERENCE STATIONARY

TRAIN TEST VALIDATION SPLIT

- How we split a dataset that is actually ordered without spoiling data structure?

1st way -

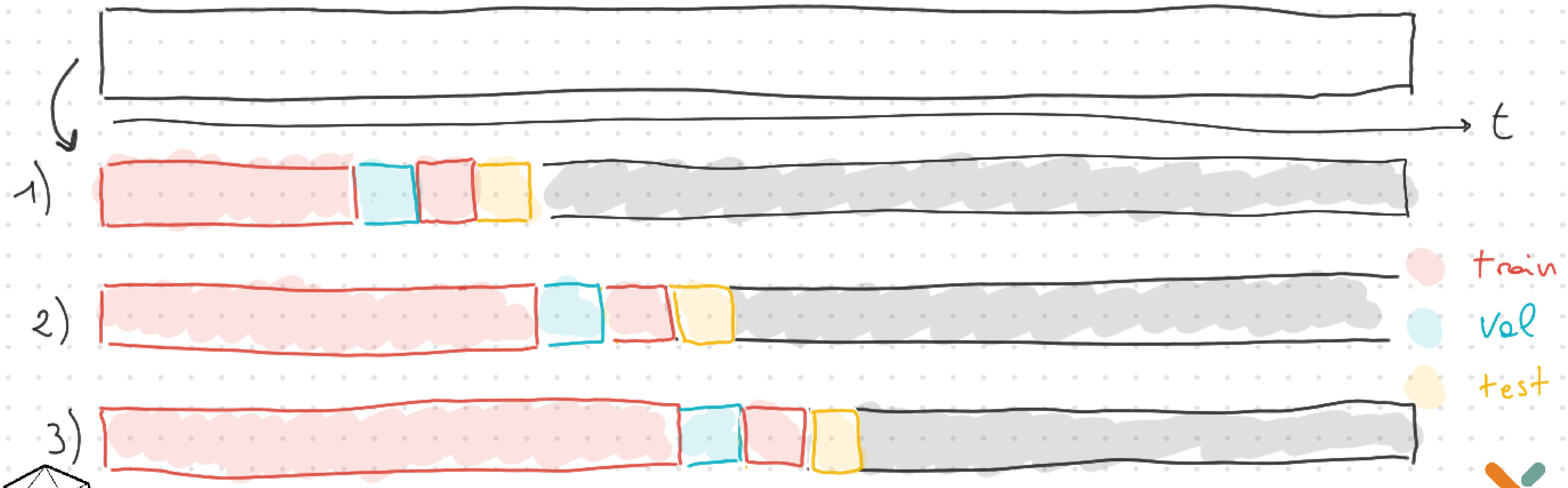


. . .

TRAIN TEST VALIDATION SPLIT

- How we split a dataset that is actually ordered without spoiling data structure?

2nd way (hv cross val)



. . .