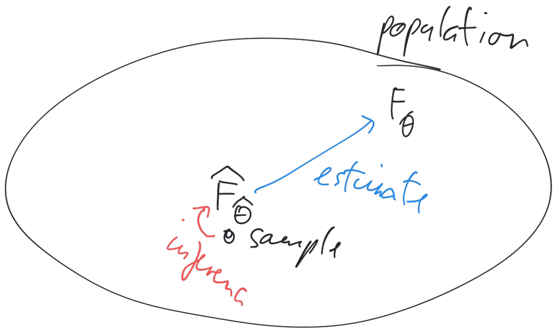
Basic concepts



infinite ensemble of TS obs.

$\{x_t^{(i)}\},\ t = 1 \dots T,\ i = 1 \dots N$

$N \to \infty$

example: $\mu_t := E_i[x_t^{(i)}]$

→ replace by a sample estimate

$\hat{\mu}_t = \bar{x} = \frac{1}{T}\sum x_t$

Ergodicity: "ensemble estim. can be replaced by sample estim."

e.g. $\underbrace{E_i[x_t^{(i)}]}_{\lim\limits_{N\to\infty} \frac{1}{N}\sum\limits_{i=1}^{N} x_t^{(i)}} = \underbrace{E_t[x_t^{(i^*)}]}_{\lim\limits_{T\to\infty} \frac{1}{T}\sum\limits_{t=1}^{T} x_t^{(i^*)}}$

Stationarity

"weak sense stat."

$E[x_t] = \mu_t = \text{const.}\ \forall t$

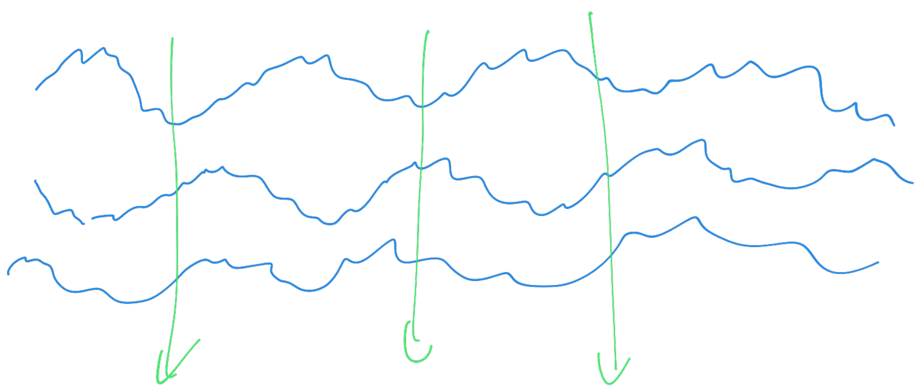
$\text{acov}[x_t, x_{t+\Delta t}] = \text{acov}[\Delta t]\ \forall t, \Delta t$

"strong stat."

$p(\{x_t \mid t_0 \le t \le t_1\}) = p(\{x_t \mid t_0 + \Delta t \le t \le t_1 + \Delta t\})$

$\forall t_0, t_1, \Delta t$

def. apply $\boxed{\text{across an ensemble of TS}}$

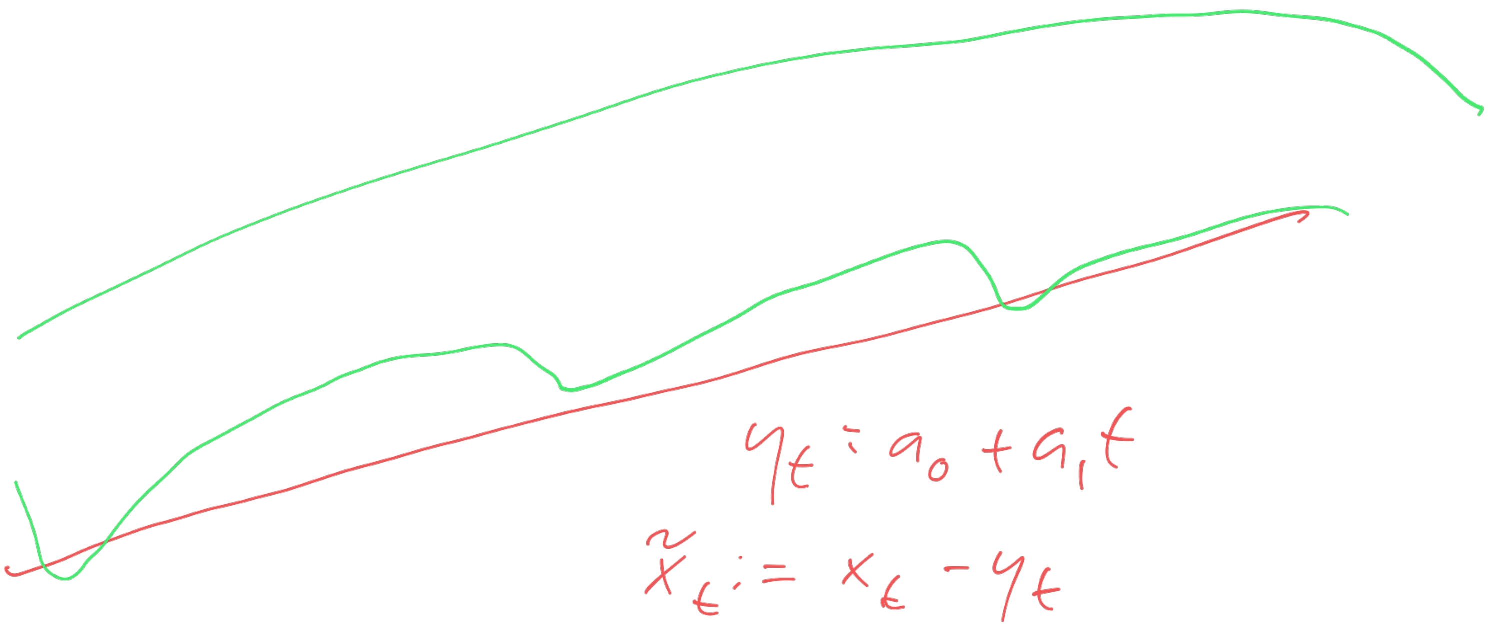


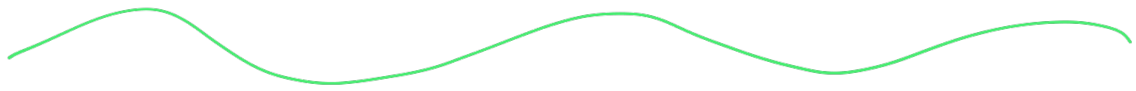
$\mu_t = \mu_{t'} = \mu_{t''}$

$x_i(t) = a \sin(\omega t + \varphi_i) + \varepsilon(t)$ ~ p

$E[x_t \mid x_{t-1}] \neq E[x_t \mid x'_{t-1}]$

for $x_{t-1} \neq x'_{t-1}$



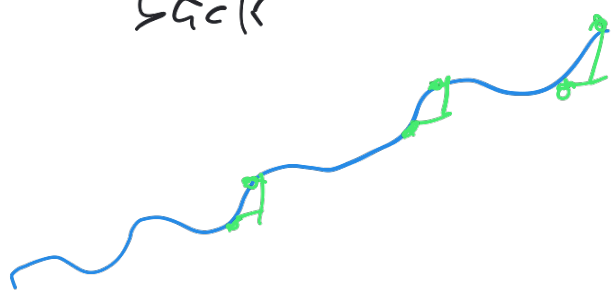


What if my TS is non-stat.?

1) power spectrum: remove lowest freq. comp. → transform back

   HPF

2) "Differencing TS"



   create new TS $\tilde{x}_t := x_t - x_{t-1}$

   → may amplify noise!

   → may remove other important aspects of TS

3) fit a model to TS that captures systematic var. → subtract off

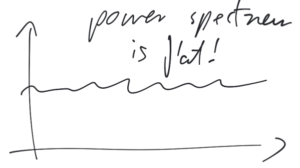
Def. "white noise"

"no corr. in TS at all"

$E[x_t] = 0$

$E[x_t x_{t'}] = \begin{cases} \sigma^2 & \text{for } t = t' \\ 0 & \text{for } t \neq t' \end{cases}$

$x_t \sim W(0, \sigma^2)$



Def. acov & acorr func.

$$\mathrm{acov}[x_t, x_{t+\Delta t}] := E[(x_t - \mu_t)(x_{t+\Delta t} - \mu_{t+\Delta t})]$$
$$= E[x_t x_{t+\Delta t}] - \mu_t \mu_{t+\Delta t} \equiv \gamma(x_t, x_{t+\Delta t})$$

$\hookrightarrow$ def. across infinite ensemble

$$\mathrm{acorr}[x_t, x_{t+\Delta t}] := \frac{\gamma(x_t, x_{t+\Delta t})}{\sigma_t \, \sigma_{t+\Delta t}}$$
$$\equiv \rho(x_t, x_{t+\Delta t}) \in [-1, +1]$$

cross-cov. among TS $\{x_t\}$ and $\{y_t\}$

$$\mathrm{xcov}[x_t, y_{t+\Delta t}] := E[(x_t - \mu_t^{(x)})(y_{t+\Delta t} - \mu_{t+\Delta t}^{(y)})]$$
$$\equiv \gamma(x_t, y_{t+\Delta t})$$
$$\mathrm{xcorr}[x_t, y_{t+\Delta t}] := \frac{\gamma(x_t, y_{t+\Delta t})}{\sigma_t^{(x)} \, \sigma_{t+\Delta t}^{(y)}}$$

Remarks

assume ergodicity and stationarity

1) $\Rightarrow$ $\mu_t = \mu_{t+\Delta t}$

$\sigma_t^2 = \sigma_{t+\Delta t}^2$ for any $\Delta t, t$

$\mathrm{acov}[x_t, x_{t+\Delta t}] = \mathrm{acov}(\Delta t)$

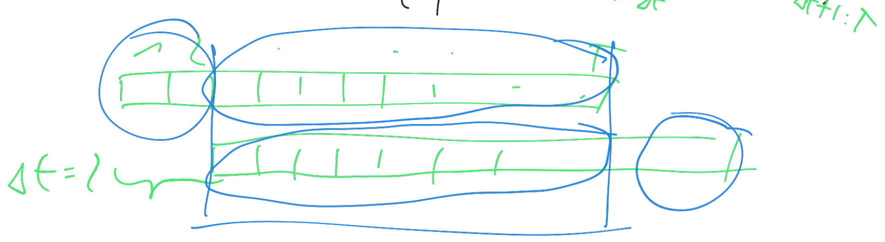
$$\rho(\Delta t) := \frac{\gamma(\Delta t)}{\gamma(0)}$$

2) replace ensemble estim. $\hookrightarrow$ sample estim.

$$\hat{\mu} = \bar{x} = \frac{1}{T} \sum_{t=1}^{T} x_t \quad , \{x_t\}, t = 1 \ldots T$$

$$\hat{\sigma}^2 = s_x^2 = \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x})^2$$

$$\hat{\gamma}(\Delta t) = \frac{1}{T - \Delta t} \sum_{t=1}^{T-\Delta t} (\underbrace{x_t}_{1:T-\Delta t} - \bar{x})(\underbrace{x_{t+\Delta t}}_{\Delta t+1:T} - \bar{x})$$



3) $\rho(\Delta t) = \rho(-\Delta t)$

4) "1:1" relation betw. acorr. func. & power spectrum

$\rightarrow$ Wiener-Khintchin theorem

assume periodic func.

$x(t) = x(t + \Delta t) \quad \forall t; \Delta t$ fixed

Auto-regressive Moving-Average (ARMA) models

- provide a baseline for RNN

obs.: $\{x_t\}, t=1\ldots T$

univariate $x_t \in \mathbb{R}$ $\quad\varepsilon_t \sim W(0,\sigma^2)$

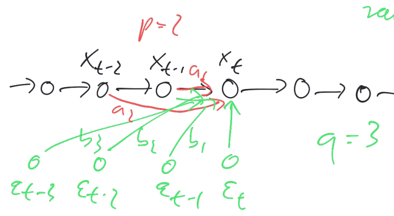
"offset"

$$x_t = a_0 + \underbrace{\sum_{i=1}^{p} a_i x_{t-i}}_{AR(p)} + \underbrace{\underbrace{\sum_{j=1}^{q} b_j \varepsilon_{t-j}}_{MA(q)} + \varepsilon_t}_{\text{random part}},$$

$$+ \sum_{k=0}^{r} c_k y_{t-k}$$

ARMAX

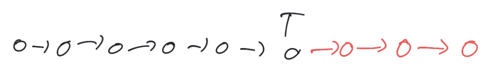
exogenous inputs



- inference (stat.)
- training (ML)
- fitting

What to do with such a model?

- Is model a good descrip. of TS?
- stationary?
- optimal order p, q? $ARMA(p,q)$
- test hyp.: $a_i \neq 0$, $b_j \neq 0$
  $a_i > \theta$ $\quad c_k \neq 0$
- prediction



ARMA(p,q)
run forward in time

Properties of ARMA models

- **Basic duality betw. AR and MA parts**

assume AR(1) model:

$$x_t = a_0 + a_1 x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim W(0,\sigma^2)$$

"initialize" $\quad x_1 = a_0 + \varepsilon_1$

$$E[x_T] = a_0 + a_1 x_{T-1} + \varepsilon_T$$

$$= a_0 + a_1(a_0 + a_1 x_{T-2} + \varepsilon_{T-1}) + \varepsilon_T$$

$$= a_0 + a_1(a_0 + a_1(a_0 + a_1 x_{T-3} + \varepsilon_{T-2}) + \varepsilon_{T-1}) + \varepsilon_T$$

$$= \ldots = E\left[a_0 \sum_{i=0}^{T-1} a_1^i + \sum_{i=0}^{T-1} a_1^i \varepsilon_{T-i}\right] \qquad E[\varepsilon] = 0$$

$$\frac{a_0}{1-a_1} \qquad \to \frac{a_0}{1-a_1} \text{ for } T \to \infty \text{ if } |a_1| < 1 \qquad 0$$

1) expanded into $MA(\infty)$, $T \to \infty$
2) for $|a_1| > 1$: will diverge!!! (non-stationary)

$$E[\Sigma a \varepsilon] = \Sigma a E[\varepsilon] = 0$$

3) $E[x_T] = \dfrac{a_0}{1-a_1} \neq a_0$ for $a_0, a_1 \neq 0$

- reexpress $MA(1)$ as an $AR(\infty)$

$$x_T = \varepsilon_T + b_1 \underbrace{\varepsilon_{T-1}}_{= x_{T-1} - b_1 \varepsilon_{T-2}} = \varepsilon_T + b_1(x_{T-1} - b_1 \varepsilon_{T-2})$$

$$= \ldots = \sum_{i=1}^{T-1} b_1^i x_{T-i} + \varepsilon_T$$

- Relation betw. coeff. of our AR process and the acov func.

- example: AR(1)

$$x_t = a_0 + a_1 x_{t-1} + \varepsilon_t \quad , \varepsilon_t \sim W(0, \sigma^2)$$

- assume $a_0 = 0$, otherwise subtract off $E[x_t] = \frac{a_0}{1-a_1}$ provided $|a_1| < 1$

$$E[x_t x_{t-1}] = E[a_1 x_{t-1} x_{t-1}] + \underbrace{E[\varepsilon_t x_{t-1}]}_{\underbrace{E[\varepsilon_t]}_{0} E[x_{t-1}] = 0}$$

$E[\varepsilon_t \varepsilon_{t'}] = 0$ for $t \neq t'$

$$x_{t-1} = \sum_{i=0}^{t-2} a_1^i \varepsilon_{t-1-i} + \varepsilon \ldots$$