Contents

1 Time-series Analysis

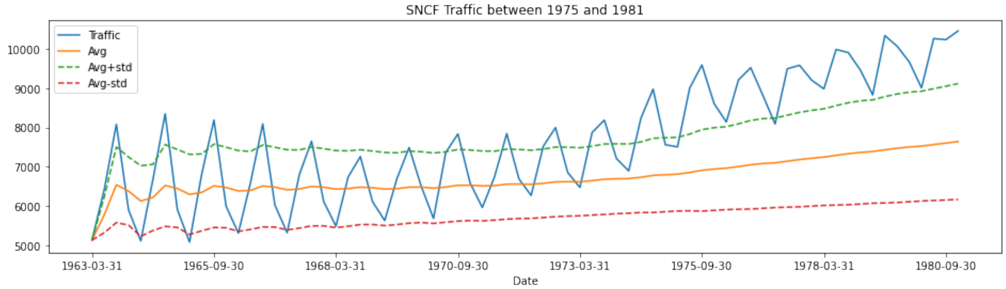
1.1 Properties of Times Series

1.1.1 Descriptive Statistics estimators

Definition (Sample mean and std).

$$\mathrm{\bar{x}_t} = \frac{1}{\mathrm{t}}\sum_{i=1}^{\mathrm{t}} \mathrm{x_i}; \quad \hat{\sigma}_t^2 = \frac{1}{t}\sum_{i=1}^{t}(x_i - \bar{\mathrm{x}}_t)^2$$

Example. To calculate the bandwidth of this time-series we iteratively evaluate the mean and std as we move across time:



Definition (Autocovariance). The covariance between two time points is referred to as the autocovariance and is defined as:

$$K_X(s,t) = \mathrm{cov}\,(X_t, X_s) = E\left[(x_s - \mu_s)(x_t - \mu_t)\right]; \quad \widehat{K}_X(h) = \frac{1}{t-h}\sum_{i=1}^{t-h}(x_i - \bar{x}_t)(x_{i+h} - \bar{x}_t)$$

- Autocorrelation. We can normalize the autocovariance to obtain; It is defined as $\hat{\rho}_t(h) \in [-1, 1]$

$$\rho_X(s,t) = \frac{K_X(s,t)}{\sqrt{K_X(s,s)K_X(t,t)}}; \quad \hat{\rho}(h) = \frac{\hat{K}_X(h)}{\hat{K}(0)}$$

Example. Let the time-series $X = [1750, 1560, 1820, 2090]$ (with $t = 4$) the first five months of the SNCF traffic in 1963.
- Autocovariance ($h = 1$). Let's look at the correlation at a lag of $h = 1$:

$$\begin{aligned}
\widehat{K}_X(1) &= \frac{1}{3}\sum_{i=1}^{3}(x_i - \bar{x}_4)(x_{i+1} - \bar{x}_4); \quad \bar{x}_4 = \frac{1}{4}(x_1 + x_2 + x_3 + x_4) = 1805 \\
&= \frac{1}{3}\Big\{(x_1 - \bar{x}_4)(x_2 - \bar{x}_4) + (x_2 - \bar{x}_4)(x_3 - \bar{x}_4) + (x_3 - \bar{x}_4)(x_4 - \bar{x}_4)\Big\} \\
&= \frac{1}{3}\Big\{13475 - 3675 + 4275\Big\} = 4691
\end{aligned}$$

And regarding $h = 2$

$$\begin{aligned}
\widehat{K}_X(2) &= \frac{1}{2}\sum_{i=1}^{2}(x_i - \bar{x}_4)(x_{i+2} - \bar{x}_4); \quad \bar{x}_4 = \frac{1}{4}(x_1 + x_2 + x_3 + x_4) = 1805 \\
&= \frac{1}{2}\Big\{(x_1 - \bar{x}_4)(x_3 - \bar{x}_4) + (x_2 - \bar{x}_4)(x_4 - \bar{x}_4)\Big\}
\end{aligned}$$

- Autocorrelation $(h=1)$**.** Then we can calculate the empirical autocorrelation

$$\hat{\rho}(1) = \frac{\hat{K}_X(1)}{\hat{K}(0)} = 0.097$$

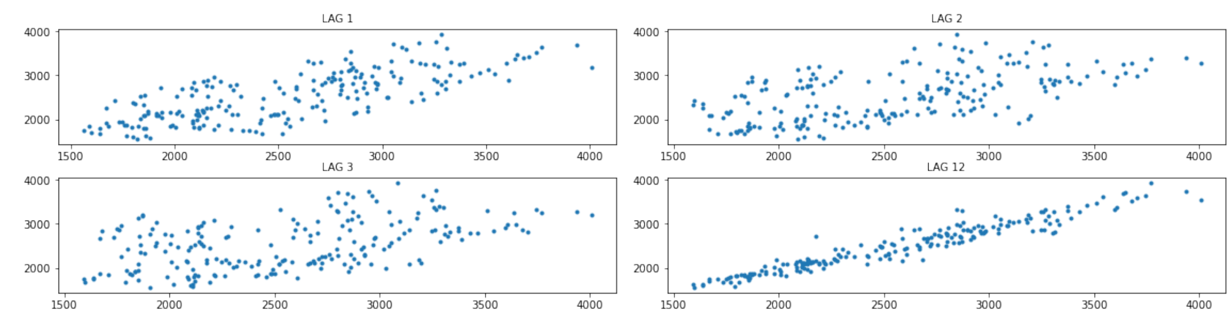
Where $\hat{K}(0) = \frac{1}{4}\Big\{(x_1 - \bar{x}_4)^2 + (x_2 - \bar{x}_4)^2 + (x_3 - \bar{x}_4)^2 + (x_4 - \bar{x}_4)^2\Big\} = 4836.82.$

- Autocorrelation matrix.

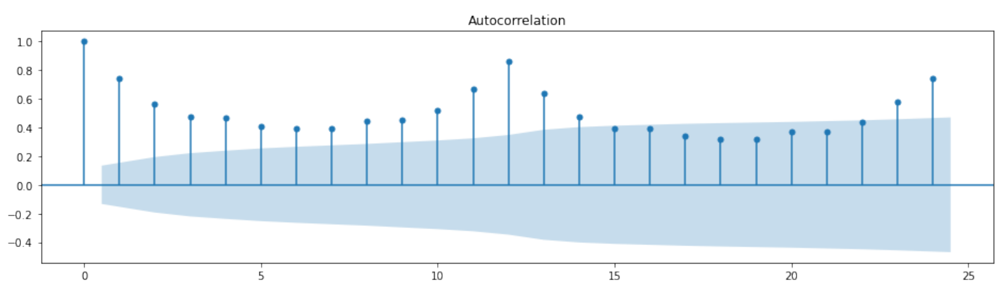
Example. Looking at our SNCF time-series we can build a lagged matrix

Date	Traffic	TMINUS1	TMINUS2	TMINUS3	TMINUS12
1963-01-01	1750	NaN	NaN	NaN	NaN
1963-02-01	1560	1750.0	NaN	NaN	NaN
1963-03-01	1820	1560.0	1750.0	NaN	NaN
1963-04-01	2090	1820.0	1560.0	1750.0	NaN
1963-05-01	1910	2090.0	1820.0	1560.0	NaN

The thinner the plot is the closer the autocorrelation is to ± 1. On the SNCF example, we note the dependency between x_t and x_{t+1} but most importantly it is significant for x_t and x_{t+12}.



Additionally there is directly the Autocorrelation plot that calculate the sample autocorrelation on the table above



Definition (Cross-Covariance).

$$K_{XY}(t,s) = \mathrm{cov}\,(X_t, Y_s)\,; \quad \widehat{K}_{XY}(h) = \frac{1}{h}\sum_{t=1}^{n-h}(x_{t+h}-\bar{x})(X_t-\bar{y})$$

- Cross-correlation.

$$\rho_{XY}(t,s) = \frac{K_{XY}(t,s)}{K_X(t,t)K_Y(s,s)}; \quad \hat{\rho}_{XY}(h) = \frac{\hat{K}_{XY}(h)}{\sqrt{\hat{K}_X(0)\hat{K}_Y(0)}}$$

Example.

t	1	2	3	4	5
X_t	1750	1560	1820	2090	1980
X_t	2010	2000	2005	1990	1995

- Crossvariance.
- Cross-correlation.
- Cross-correlation matrix.

1.1.2 White noise

White noise must have uniform distribution over frequencies but it can have any distribution over time eg

$$w_t, t = 1, .., T, \quad w_t \sim N(0, \sigma^2)$$

1.1.3 Difference btw statistics and time-series

in stat it is iid and time-series it is dependent on the past.

One of the implications is that we cannot use linear regression in this setting since MLE implies that the samples are iid.

1.1.4 Stationary processes

1.1.5 Test for Stationary and Autocorrelation

1.1.5.1 Dickey Fuller Test

A big deal in time series modeling is to make sure the time serie is stationary. Of course we need some robust test and that is the Dickey Fuller Test.

Dickey Fuller test. It assumes our times series is a AR(1) process that is

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$$

The hypothesis H0 means the time series has a unit root and ¡1 has not unit root.

$$\begin{array}{ll} H_0 & \phi_1 = 1 \\ H_1 & \phi_1 < 1 \end{array}$$

The first things we do is substracting

$$y_t - y_{t-1} = \mu + (\phi_1 - 1)\, y_{t-1} + \varepsilon_t$$

We introduce the notation

$$\Delta y_t = \mu + \delta y_{t-1} + \varepsilon_t$$

Delta is phi1 -1. This transformation we can write it

$$\begin{array}{l} H_0 \quad \delta = 0 \\ H_1 : \delta < 0 \end{array}$$

The intent of doing this transformation is that?? We cannot apply the t-test cause y_{t-1} is still stationary cause X_t is assumes non stationary under the null hypothesis. But it turns out we can compute the exact t-stat but the distribution is not the t-distribution but the dickey-faller distribution. Explicityly

$$t_{\hat{\delta}} = \frac{\hat{\delta}}{\mathrm{se}(\hat{\delta})}$$

It is simialr to the t-stat but instead to compare the t-stat against the normal distribution we do it to the dickey fuller distribution.

$$t_{\hat{\delta}} < DF_{\text{critical}} \;\rightarrow\; \text{Reject } H_0$$

We reject the fact that it has a unit root so stationary

$$t_{\hat{\delta}} > DF_{\text{critical}} \;\rightarrow\; \text{Don't reject } H_0$$

1.1.5.2 Augmented Dickey Fuller Test

Of course most time series will be more complex than an AR(1) that's where this test shows ups. Here it is a more complicated AR(p) model

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t$$

We do the same transformation

$$\Delta X_t = \mu + \delta y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \varepsilon_t$$

We do the same test

$$\begin{array}{ll} H_0 & \delta = 0 \\ H_1 & \delta < 0 \end{array}$$

And we get where we can apply a t-test on those

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i}{\mathrm{se}\left(\hat{\beta}_i\right)}$$

1.2 ARIMA Models for Times Series

1.2.1 Autoregressive process $AR(p)$

1.2.1.1 Definition

Definition (AR(p)). The AR process is a way to encode into the white noise process how past influence the future. The formula is

$$\boxed{X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + w_t; \quad w_t \sim \mathcal{N}(0, \sigma^2)}$$

Example. If I let $p = 1, \phi_1 = 1$ we obtain the system

$$X_t = X_{t-1} + w_t \iff \begin{cases} X_1 = X_0 + w_1 \\ X_2 = X_1 + w_2 \end{cases}$$

Alternatives. If we let w_t be a binary noise instead of a gaussian noise with equal probability we obtain a <u>symmetric arithmetic random walk</u> $AR(1)$:

$$X_t = X_{t-1} + w_t$$

1.2.1.2 Estimation for AR processes of the lags p (ACF and PACF)

(i) Autocorrelation function (ACF). We consider an AR(1) process

$$X_t = \phi X_{t-1} + w_t, \quad |\phi| < 1$$

The condition on ϕ makes the T.S. stationary with zero mean. We can rewrite with lag operators

$$(1 - \phi L) X_t = \varepsilon_t$$

We multiply on both sides with $(1 - \phi L)^{-1}$ to obtain

$$\begin{aligned} (1 - \phi L)^{-1}(1 - \phi L) X_t &= (1 - \phi L)^{-1} \varepsilon_t \\ X_t &= (1 - \phi L)^{-1} \varepsilon_t \end{aligned}$$

Then using the result (geometric rule) $(1 - z)^{-1} = 1 + z + z^2 + z^3 + \ldots$ which holds for $|z| < 1$we get the

$$(1 - \phi L)^{-1} = \lim_{j \to \infty} \left(1 + \phi L + (\phi L)^2 + \ldots + (\phi L)^j\right)$$

We get the updated formula:

$$X_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \phi^3 \varepsilon_{t-3} + \ldots = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

- Moments. The variance is then,

$$\begin{aligned} K(0) = \operatorname{var}[X_t] &= \mathbb{E}[X_t - \mathbb{E}[X_t]]^2 \\ &= \mathbb{E}\left[\varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \phi^3 \varepsilon_{t-3} + \ldots\right]^2 \\ &= \operatorname{var}[\varepsilon_t] + \phi^2 \operatorname{var}[\varepsilon_{t-1}] + \phi^4 \operatorname{var}[\varepsilon_{t-2}] + \phi^6 \operatorname{var}[\varepsilon_{t-3}] + \ldots \\ &= \left(1 + \phi^2 + \phi^4 + \phi^6 + \ldots\right) \sigma^2 \\ &= \frac{1}{1 - \phi^2} \sigma^2 \end{aligned}$$

The first order covariance is

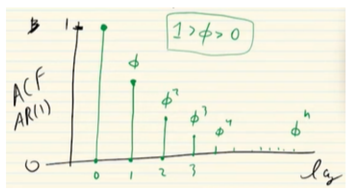
$$
\begin{aligned}
K_X(1) &= \mathbb{E}\Big[\big(y_t - \mathbb{E}[y_t]\big)\big(y_{t-1} - \mathbb{E}[y_{t-1}]\big)\Big] \\
&= \mathbb{E}\left[(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \ldots)(\varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \phi^3\varepsilon_{t-4} + \ldots)\right] \\
&= \left(\phi + \phi^3 + \phi^5 + \ldots\right)\sigma^2 = \phi\left(1 + \phi^2 + \phi^4 + \phi^6 + \ldots\right)\sigma^2 \\
&= \phi\frac{1}{1-\phi^2}\sigma^2 \\
&= \phi\mathsf{var}\left[X_t\right] \qquad (1)
\end{aligned}
$$

while for $j > 1$ we have,

$$
K_X(j) = \mathbb{E}\Big[\big(X_t - \mathbb{E}[X_t]\big)\big(X_{t-j} - E[y-j]\big)\Big] = \phi^j\mathsf{var}\left[X_t\right] \qquad (2.13)
$$

As a consequence the coefficients can be written as

$$
\rho(0) = \frac{K(0)}{K(0)} = 1,\ \ \rho(1) = \frac{K_X(1)}{K(0)} = \phi,\ \ \ldots\ \ ,\ \ \rho(j) = \frac{K_X(j)}{K(0)} = \phi^j \qquad (2.14)
$$



(ii) Partial Autocorrelation function. Consider the simpler AR(1) process defined as $X_t = \phi X_{t-1} + w_t$. The PACF is defined as

$$
\varphi_X(h) = \operatorname{corr}\left(X_t - \hat{X}_t, X_{t-h} - \hat{X}_{t-h}\right)
$$

To properly consider the partial autocorrelation, we need to compute the least squares estimator $\hat{X}_t$ for X_t based on previous time points. For example, to compute $\varphi(2)$, we take $\hat{X}_t = \hat{\beta}X_{t-1}$ where $\hat{\beta}$ is chosen to minimize the least square equation:

$$
\begin{aligned}
&\mathrm{E}\left(X_t - \hat{X}_t\right)^2 = \mathrm{E}\left(X_t - \beta X_{t-1}\right)^2 \\
&= \mathrm{E}\left(X_t^2\right) - 2\beta\mathrm{E}\left(X_t X_{t-1}\right) + \beta^2\mathrm{E}\left(X_{t-1}^2\right) = \left(1+\beta^2\right)K_X(0) - 2\beta K_X(1)
\end{aligned}
$$

By taking the derivative with respect to β, we can find the critical point $\hat{\beta} = K_X(1)/K_X(0)$. Similarly, for $\hat{X}_{t-2} = \hat{\beta}X_{t-1}$, we have

$$
\begin{aligned}
&\mathrm{E}\left(X_{t-2} - \beta X_{t-1}\right)^2 \\
&= \mathrm{E}\left(X_{t-2}^2\right) - 2\beta\mathrm{E}\left(X_{t-2}X_{t-1}\right) + \beta^2\mathrm{E}\left(X_{t-1}^2\right) = \left(1+\beta^2\right)K_X(0) - 2\beta K_X(1)
\end{aligned}
$$

For the general AR(p) process, $X_t = w_t + \sum_{i=1}^{p}\phi_i X_{t-i}$, we have a similar set up. For lags $h > p$, if we assume for now that the least squares estimator is

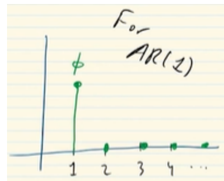
$$
\hat{X}_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p},
$$

then we get a similar calculation as above. Namely that

$$
\operatorname{corr}\left(X_t - \hat{X}_t, X_{t-h} - \hat{X}_{t-h}\right) = \operatorname{corr}\left(w_t, X_{t-h} - \hat{X}_{t-h}\right) = 0
$$

In the case that the lag is less than or equal to p, we need to determine how the estimate the coefficients β_i before computing the PACF.

$$\varphi_X(h) = \begin{cases} \phi, & h = 1 \\ 0, & h > 1 \end{cases}$$



1.2.1.3 Estimation for AR processes of the $\hat{\phi}_i$

There are two approaches to estimate the parameters of the $AR(p)$ process the $(\hat{\phi}_1, \hat{\phi}_2, .., \hat{\phi}_p)$ that we will consider:

1. The Yule-Walker equation
2. The MLE estimator

(i) Yule-Walker equation. Let a sample set $\{X_1, .., X_T\}$
- Step 1. We estimate the mean

$$\bar{X} = \frac{1}{T} \sum_{t=1}^{T} X_t$$

And then we consider the centered time series $X_i - \bar{X}$ aka we now assume that $\mathbb{E}[X_t] = 0$.

- Step 2. Estimate the variance σ^2 of w_t

$$\begin{aligned} X_t &= w_t + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} \\ \mathbb{E}[X_t X_t] &= \mathbb{E}[(w_t + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p}) X_t] \quad \text{Add E[] and Xt on both sides} \\ K_X(0) = cov(X_t, X_t) = Var(X_t) &= cov(w_t, X_t) + \phi_1 cov(X_{t-1}, X_t) + .. + \phi_p cov(X_{t-p}, X_t) \\ K_X(0) &= \sigma^2 + \phi_1 K_X(1) + .. + \phi_p K_X(p) \\ \sigma^2 &= K_X(0) - \phi_1 K_X(1) - \ldots - \phi_p K_X(p) \end{aligned}$$

Note we used the fact that $cov(w_t, X_t) = cov(w_t, w_t)$ and the equality

$$cov(X, Y) = \mathbb{E}(X - \mu_X)(Y - \mu_Y)$$

Where recollect that both r.v. are centered as such $\mu_X = \mu_Y = 0$. Hence, we can use the estimates for the autocovariance to estimate σ^2.

$$\begin{aligned} \hat{\sigma}^2 &= \hat{K}_X(0) - \phi_1 \hat{K}_X(1) - \ldots - \hat{\phi}_p K_X(p) \\ &= \hat{K}_X(0) - \sum_{i=1}^{p} \hat{\phi}_i \hat{K}_X(i) \end{aligned}$$

The problem is that the $\hat{\phi}_i$ are unknown.

- Step 3. Estimate the $\phi' s$; We can consider more equations based on the autocovariance at lags 1 through p.

$$
\begin{aligned}
& K_X(1)=\operatorname{cov}\left(X_{t-1}, X_t\right)=\phi_1 K_X(0)+\phi_2 K_X(1)+\ldots+\phi_p K_X(p-1) \\
& K_X(2)=\operatorname{cov}\left(X_{t-2}, X_t\right)=\phi_1 K_X(1)+\phi_2 K_X(0)+\ldots+\phi_p K_X(p-2) \\
& \vdots \\
& K_X(p)=\operatorname{cov}\left(X_{t-p}, X_t\right)=\phi_1 K_X(p-1)+\phi_2 K_X(p-2)+\ldots+\phi_p K_X(0)
\end{aligned}
$$

Here, we have p linear equations with p unknowns, which can be written as $K=\Gamma \phi$ where

$$
K=\left(\begin{array}{c}
K_X(1) \\
\vdots \\
K_X(p)
\end{array}\right), \quad \phi=\left(\begin{array}{c}
\phi_1 \\
\vdots \\
\phi_p
\end{array}\right), \quad \Gamma=\left(\begin{array}{cccc}
K_X(0) & K_X(1) & \cdots & K_X(p-1) \\
K_X(1) & K_X(0) & \cdots & K_X(p-2) \\
\vdots & \ddots & \ddots & \vdots \\
K_X(p-1) & K_X(p-2) & \cdots & K_X(0)
\end{array}\right)_{p \times p}
$$

We can also write $\sigma^2=K_X(0)-\phi^{\mathrm{T}} K$. This system of $p+1$ equations is known as the Yule-Walker Equations. We can solve for the coefficients $\phi=\Gamma^{-1} K$ as the matrix Γ is positive definite. As a result we can solve step 2 with $\sigma^2=K_X(0)-K^{\mathrm{T}} \Gamma^{-1} K$, because Γ is symmetric.

We can use the estimator for the autocovariance from Chapter 1 to get a data driven estimate for the parameters for this time series:

$$
\hat{\phi}=\hat{\Gamma}^{-1} \hat{K}, \quad \text { and } \quad \hat{\sigma}^2=\hat{K}_X(0)-\hat{K}^{\mathrm{T}} \hat{\Gamma}^{-1} K
$$

(ii) Conditional MLE. Given that $w_t \sim N(0, \sigma^2)$. Beginning with the causal $AR(1)$ process $X_t=\mu+\phi(X_{t-1}-\mu)+w_t$ we use the recursive definition of the process to write the likelihood as a joint probability $f(\cdot)$:

$$
\begin{aligned}
L\left(\mu, \phi, \sigma^2 ; X_1, \ldots, X_T\right) & =f\left(X_1, \ldots, X_T ; \mu, \phi, \sigma^2\right) \\
& =f\left(X_1\right) f\left(X_2 \mid X_1\right) \ldots f\left(X_T \mid X_{T-1}, .., X_1\right) \\
& =f\left(X_1\right) f\left(X_2 \mid X_1\right) \ldots f\left(X_T \mid X_{T-1}\right) \quad \text{Cause Markov Process} \\
& =f\left(X_1\right) \prod_{t=2}^T p(X_t \mid X_{t-1})
\end{aligned}
$$

- Step 1. We have demonstrated previously that

$$
X_1 \sim \mathcal{N}\left(\mu, \sigma^2 /\left(1-\phi^2\right)\right)
$$

- Step 2. The conditional distribution is

$$
X_t \mid X_{t-1} \sim \mathcal{N}\left(\mu+\phi\left(X_{t-1}-\mu\right), \sigma^2\right)
$$

Since

$$
\begin{aligned}
X_t & =w_t+\phi(X_{t-1}-\mu)+\mu \\
X_t \mid X_{t-1} & =w_t+\phi(X_{t-1}-\mu)+\mu
\end{aligned}
$$

By conditioning X_{t-1} becomes a constant and he is no longer a r.v. Yet $w \sim \mathcal{N}(0, \sigma^2)$ remains one.

- Step 3. Then we put it all together to obtain the likelihood:

$$
\begin{aligned}
L\left(\mu, \phi, \sigma^2\right)= & \frac{\left(1-\phi^2\right)^{1 / 2}}{\left(2 \pi \sigma^2\right)^{T / 2}} \\
& \times \exp \left[-\frac{1}{2 \sigma^2}\left\{\left(X_1-\mu\right)^2\left(1-\phi^2\right)+\sum_{t=2}^T\left(\left(X_t-\mu\right)-\phi\left(X_{t-1}-\mu\right)\right)^2\right\}\right]
\end{aligned}
$$

Next to obtain the parameters

$$
\begin{aligned}
\frac{\partial \log(L)}{\partial \sigma^2} = 0 \iff \hat{\sigma}^2 &= S_u(\mu, \phi)/T \\
\hat{\phi} &= \frac{\sum_{t=2}^{T} \left(X_t - \bar{X}_{(2)}\right)\left(X_{t-1} - \bar{X}_{(1)}\right)}{\sum_{t=2}^{T} \left(X_{t-1} - \bar{X}_{(1)}\right)^2} \\
\hat{\mu} &= \frac{\bar{X}_{(2)} - \hat{\phi}\bar{X}_{(1)}}{1 - \hat{\phi}}
\end{aligned}
$$

1.2.1.4 Forecasting

Another significant problem in time series analysis is that of forecasting. That is, given data $X_1, \ldots, X_T$, we want to compute the best predictions for subsequent time points $X_{T+1}, X_{T+2}, \ldots, X_{T+m}$. We won't initially assume that the process is autoregressive, but we will assume that X_t is stationary. First, we can consider linear predictors, which are those of the form

$$
\hat{X}_{T+m} = a_0 + \sum_{t=1}^{T} a_t X_t
$$

where we want to make a good choice of parameters a_t. To do that, we minimize the squared error as usual:

$$
\mathscr{L} = \underset{a_0, a_1, \ldots, a_T}{\arg\min} \; \mathrm{E}\left\{\left(X_{T+m} - a_0 - \sum_{t=1}^{T} a_t X_t\right)^2\right\}.
$$

Taking the derivative with respect to each a_t gives a system of equations (and using the chain rule $f(g(x))' = f'(g(x))g'(x)$):

$$
\begin{aligned}
\frac{\partial \mathscr{L}}{\partial a_0} = 0 \iff 0 =& \mathrm{E}\left(X_{T+m} - \hat{X}_{T+m}\right) \\
\frac{\partial \mathscr{L}}{\partial a_1} = 0 \iff 0 =& \mathrm{E}\left(\left(X_{T+m} - \hat{X}_{T+m}\right) X_1\right) \\
&\vdots \\
\frac{\partial \mathscr{L}}{\partial a_T} = 0 \iff 0 =& \mathrm{E}\left(\left(X_{T+m} - \hat{X}_{T+m}\right) X_T\right)
\end{aligned}
$$

- Estimate of a_0. Let the mean of the process be μ. By the first equation, (and since we have a stationary process):

$$
\begin{aligned}
\mu = \mathrm{E}\left(X_{T+m}\right) = \mathrm{E}\left(\hat{X}_{T+m}\right) = \mathrm{E}\left(a_0 + \sum_{t=1}^{T} a_t X_t\right) = a_0 + \sum_{t=1}^{T} a_t \mu \\
a_0 = \mu\left(1 - \sum_{t=1}^{T} a_t\right)
\end{aligned}
$$

Hence, we have $\hat{X}_{t+m} = \mu + \sum_{t=1}^{T} \alpha_t \left(X_t - \mu\right)$. Thus, we can centre the process and consider time series with $\mu = 0$ and $a_0 = 0$.

- Estimates of the a_i. For a one-step-ahead prediction, which is to estimate $\hat{X}_{T+1}$, we solve the above equations to get

$$
\begin{aligned}
0 =& \mathrm{E}\left(\left(X_{T+1}-\hat{X}_{T+1}\right) X_1\right) = K_X(T) - \sum_{t=1}^{T} a_t K_X(t-1) \\
0 =& \mathrm{E}\left(\left(X_{T+1}-\hat{X}_{T+1}\right) X_2\right) = K_X(T-1) - \sum_{t=1}^{T} a_t K_X(t-2) \\
&\vdots \\
0 =& \mathrm{E}\left(\left(X_{T+m}-\hat{X}_{T+m}\right) X_T\right) = K_X(1) - \sum_{t=1}^{T} a_t K_X(T-t)
\end{aligned}
$$

If similar to before we let K be the T-long vector with entries $K_X(T), \ldots, K_X(1)$

$$
\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} K_X(T) \\ \vdots \\ K_X(1) \end{pmatrix} - \begin{pmatrix} K_X(0) & K_X(1) & \cdots & K_X(T-1) \\ K_X(1) & K_X(0) & \cdots & K_X(T-2) \\ \vdots & \ddots & \ddots & \vdots \\ K_X(T-1) & K_X(T-2) & \cdots & K_X(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_T \end{pmatrix}
$$

and let

$$
K = \begin{pmatrix} K_X(T) \\ \vdots \\ K_X(1) \end{pmatrix} \quad \Gamma = \begin{pmatrix} K_X(0) & K_X(1) & \cdots & K_X(T-1) \\ K_X(1) & K_X(0) & \cdots & K_X(T-2) \\ \vdots & \ddots & \ddots & \vdots \\ K_X(T-1) & K_X(T-2) & \cdots & K_X(0) \end{pmatrix}_{T \times T} \quad a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_T \end{pmatrix}
$$

Then, the above equations can be written as

$$
K = \Gamma a \quad \text{or} \quad a = \Gamma^{-1} K
$$

in the case that the inverse exists. And note that this equation is exactly like the Yule-Walker equation but in YW $[\Gamma]_{p\times p}$ whereas this time $[\Gamma]_{T\times T}$. As a consequence here we might not be motivated to invert the matrix cause too big. And we have to estimate $\hat{K}_X$ which requires too much lags involving a lot of numerical errors.

The solution is to let $X = (X_1, \ldots, X_T)^{\mathrm{T}}$, our one-step prediction can be written as

$$
\hat{X}_{T+1} = a^{\mathrm{T}} X = K^{\mathrm{T}} \Gamma^{-1} X
$$

As like estimation for the AR process with the Yule-Walker equations, our prediction is based on the autocovariances. If we knew what the autocovariance is-i.e. we use K and Γ instead of $\hat{K}$ and $\hat{\Gamma}$-then the mean squared prediction error is

$$
\begin{aligned}
\mathrm{E}\left(X_{T+1}-\hat{X}_{T+1}\right)^2 &= \mathrm{E}\left(X_{T+1} - K^{\mathrm{T}}\Gamma^{-1}X\right)^2 \\
&= \mathrm{E}\left(X_{T+1}^2 - 2K^{\mathrm{T}}\Gamma^{-1}XX_{T+1} + K^{\mathrm{T}}\Gamma^{-1}XX^{\mathrm{T}}\Gamma^{-1}K\right) \\
&= K_X(0) - 2K^{\mathrm{T}}\Gamma^{-1}K + K^{\mathrm{T}}\Gamma^{-1}\Gamma\Gamma^{-1}K \\
&= K_X(0) - K^{\mathrm{T}}\Gamma^{-1}K.
\end{aligned}
$$

- Forecasting for an AR(p) process. For the AR(p) process,

$$
X_t = w_t + \sum_{i=1}^{p} \phi_i X_{t-i},
$$

the one-step-ahead prediction comes precisely from estimating the coefficents as in the Yule-Walker equations to get

$$\hat{X}_{T+1} = \sum_{i=1}^{p} \hat{\phi}_i X_{T+1-i}.$$

However, if we do not know a priori that we have an order- p process, then we would have to estimate α_i for all $i = 1, \ldots, T$, which could require the inversion of a very large matrix. Thus, for general ARMA models, we have to work harder.

1.2.2 Moving Average process $MA(q)$

1.2.2.1 Definition

Definition (MA(q)). It is a *smoother* type of nosie than the white noise process. It can be expressed by the formula:

$$\boxed{X_t = \sum_{j=1}^{q} \theta_j w_{t-j} + w_t; \quad \phi_j \in \mathbb{R}}$$

Example. If we let $q = 1, \phi_1 = 1$

$$\begin{cases} X_1 = w_0 + w_1 \\ X_2 = w_1 + w_2 \\ X_3 = w_2 + w_3 \end{cases}$$

1.2.2.2 Estimation for MA processes of the lags q (ACF and PACF)

(i) Autocorrelation function (ACF). Recollect the MA(1) process

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

To evaluate the ACF we need

$$\begin{aligned} \operatorname{var}[X_t] &= \mathbb{E}[X_t - \mathbb{E}[X_t]]^2 \\ &= \mathbb{E}[(\mu + \varepsilon_t + \theta\varepsilon_{t-1}) - \mu]^2 \\ &= \mathbb{E}[\varepsilon_t]^2 + 2\theta\mathbb{E}[\varepsilon_t\varepsilon_{t-1}] + \mathbb{E}[\theta\varepsilon_{t-1}]^2 \\ &= \sigma^2 + 0 + \theta\sigma^2 \\ &= \left(1+\theta^2\right)\sigma^2 \end{aligned}$$

where $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}\left[(\varepsilon_t)^2\right] = \sigma^2$. Once again, this value is constant and does not depend on time. The first covariance may then be calculated as,

$$\begin{aligned} \mathsf{cov}[X_t, X_{t-1}] &= \mathbb{E}\Big[\big(X_t - \mathbb{E}[X_t]\big)\big(X_{t-1} - \mathbb{E}[X_{t-1}]\big)\Big] \\ &= \mathbb{E}\big[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})\big] \\ &= \mathbb{E}\left[\varepsilon_t\varepsilon_{t-1}\right] + \theta\mathbb{E}[\varepsilon_{t-1}^2] + \mathbb{E}[\theta\varepsilon_t\varepsilon_{t-2}] + \mathbb{E}[\theta^2\varepsilon_{t-1}\varepsilon_{t-2}] \\ &= 0 + \theta\sigma^2 + 0 + 0 \\ &= \theta\sigma^2 \end{aligned} \tag{2}$$

And in the more general case where we have a lag order of j,

$$\begin{aligned} \operatorname{cov}[X_t, X_{t-j}] &= \mathbb{E}[(X_t - \mathbb{E}[X_t])(X_{t-j} - \mathbb{E}[X_{t-j}])] \\ &= \mathbb{E}[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-j} + \theta\varepsilon_{t-j})] \\ &= 0 \quad \text{for } j > 1 \end{aligned}$$

The autocorrelation function (ACF) for a MA(1) process may then be derived from the expression,

$$\rho(j) \equiv \frac{K_X(j)}{K(0)}.$$

In this case $K(0)$ would refer to the variance and $K_X(j)$ would refer to the covariance for lag j. This would imply that,

$$\begin{aligned}\rho(1) &= \frac{\theta\sigma^2}{(1+\theta^2)\,\sigma^2} = \frac{\theta}{(1+\theta^2)}\\ \rho(j) &= 0 \quad \text{for} \quad j>1\end{aligned}$$

We see that for lag orders $j > 1$, the autocorrelations are zero. Hence, the autocorrelation function for a MA(1) process will go very quickly to zero as j becomes large.

(ii) Partial Autocorrelation function (PACF). For the invertible MA(1) model $X_t = w_t + \theta w_{t-1}$, which is with $|\theta| < 1$, we can write it as a convergent infinite series

$$X_t = w_t + \sum_{i=1}^{\infty} \theta^i X_{t-i}$$

in terms of the X_{t-i}. Then, applying similar tricks as above gives a least squares estimator for $\hat{X}_t = \hat{\beta} X_{t-1}$ to be $\hat{\beta} = K_X(1)/K_X(0)$. In the case of the MA(1) process, we have $\hat{\beta} = \theta/\left(1+\theta^2\right)$. Hence,

$$\begin{aligned}&\operatorname{cov}\left(X_t - \frac{\theta X_{t-1}}{1+\theta^2}, X_{t-2} - \frac{\theta X_{t-1}}{1+\theta^2}\right)\\ &\qquad = K_X(2) - \frac{2\theta}{1+\theta^2} K_X(1) + \left(\frac{\theta}{1+\theta^2}\right)^2 K_X(0) = \frac{-\theta^2}{1+\theta^2}.\end{aligned}$$

Also, the variance is

$$\begin{aligned}\operatorname{Var}\left(X_t - \frac{\theta X_{t-1}}{1+\theta^2}\right) &= K_X(0)(1+\left(\frac{\theta}{1+\theta^2}\right)^2\Bigg) - \frac{2\theta}{1+\theta^2} K_X(1)\\ &= 1+\theta^2 + \frac{\theta^2}{1+\theta^2} - \frac{2\theta^2}{1+\theta^2} = \frac{1+\theta^2+\theta^4}{1+\theta^2}\end{aligned}$$

Thus, the partial autocorrelation at lag 1 is $\varphi(1) = -\theta^2/\left(1+\theta^2+\theta^4\right)$. This can be extended to lags greater than one to show that the partial autocorrelation for the MA(1) process decreases but does not vanish as the lag increases.

1.2.3 Autoregressive Moving Average $ARMA(p, q)$

1.2.3.1 Definition

Definition (Autoregressive Moving Average Process). The time series X_t is an $ARMA(p, q)$ process if X_t has zero mean and if we can write it as

$$X_t = w_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j w_{t-j}$$

where w_t is white noise with variance σ^2 and $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q \in \mathbb{R}$ are constants with $\phi_p \neq 0$ and $\theta_q \neq 0$. Using the backshift operator B, we can succinctly write this process as $\Phi(B)X_t = \Theta(B)w_t$ where as before

$$\Phi(B) = \left(1 + \sum_{i=1}^{p} \phi_i B^i\right), \quad \text{and} \quad \Theta(B) = \left(1 + \sum_{j=1}^{q} \theta_j B^j\right)$$

1.2.4 Seasonal Autoregressive Moving Average process $SARIMA(p, q)$

1.2.4.1 Definition

1.3 Exponential smoothing (Lissage exponentiel)

1.4 Spectral density

2 Foundations

2.1 Autocorrelation plot (ACF) and PACF

Definition (Sample Autocovariance).

$$\hat{\sigma}_t^2(h) = \frac{1}{t-h} \sum_{i=1}^{t-h} (x_i - \bar{x}_t)(x_{i+h} - \bar{x}_t)$$

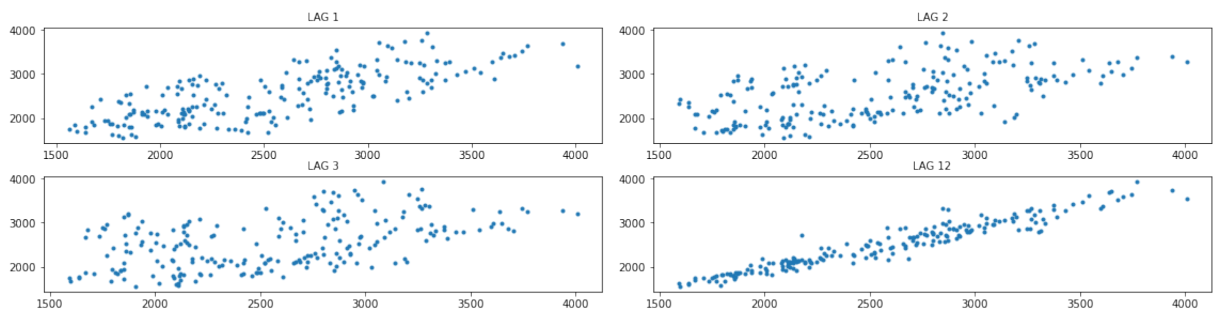
Definition (Sample Auto-correlation)

$$\hat{\rho}_t(\text{ h}) = \frac{\hat{\sigma}_1^2(\text{ h})}{\hat{\sigma}_t^2}$$

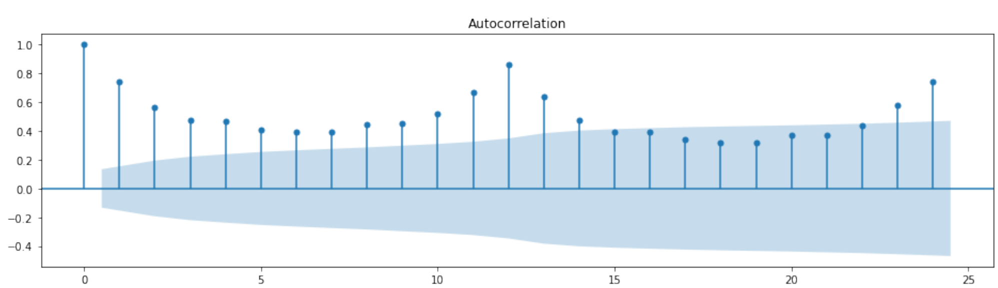
Example. It is possible to have a first idea of the autocorrelation by plotting the scatter plot of the series $\{x_t\}$ and $\{x_{t+h}\}$. We plot the couples $(Traffic, TMINUS1), (Traffic, TMINUS2), (Traffic, TMINUS3)$ and $(Traffic, TMINUS12)$

Date	Traffic	TMINUS1	TMINUS2	TMINUS3	TMINUS12
1963-01-01	1750	NaN	NaN	NaN	NaN
1963-02-01	1560	1750.0	NaN	NaN	NaN
1963-03-01	1820	1560.0	1750.0	NaN	NaN
1963-04-01	2090	1820.0	1560.0	1750.0	NaN
1963-05-01	1910	2090.0	1820.0	1560.0	NaN

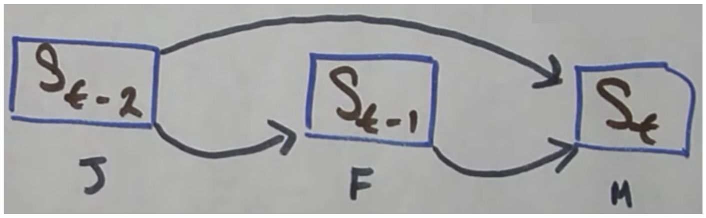
The thinner the plot is the closer the autocorrelation is to ± 1. On the SNCF example, we note the dependency between x_t and x_{t+1} but most importantly it is significant for x_t and x_{t+12}.



Additionally there is directly the Autocorrelation plot that calculate the sample autocorrelation on the table above



PACF. The difference between the Autocorrelation plot and the Partial autocorrelation plot can be though as causal relationship between three states



Let those three states be the average prices of salmon for this month, the last month and the month before last. We argue that the Autocorrelation of S_t is similar to considering the direct and indirect impact on S_t that is S_{t-2} and S_{t-1}. In contrast Partial autocorrelation is solely impacted by the direct relation S_{t-2}.

PACF tells us getting rid of the indirect effect, what is the direct effect of the price of salmon some days ago to today. ACF in contrast find all the direct and indirect effect. How to find PACF ? Let's say we are after PACF of $k = 2$

$$S_t = \phi_{21}S_{t-1} + \phi_{22}S_{t-2} + \varepsilon_t$$

We write a regression function where the coefficient ϕ_{22} will yields the direct effect of salmon two months ago on the price of salmon today. If we want to build the PACF where $k = 3$

$$S_t = \phi_{21}S_{t-1} + \phi_{22}S_{t-2} + \phi_{23}S_{t-3} + \varepsilon_t$$

Where ϕ_{23} is the direct effect.

Graphically the red bars are error bars you can think of them are when inside the error bar we do not have any evidence that this is not zero.

Figure 1: PACF for lags 2 (k=2), basically what we look at if the price of salmon two month ago impact today

Typically if we have a negative value it means that the price of salmon adversely affect at t-4 lags the price of today.

A good model would (this is similar to an autoregressive model)

$$S_t = \beta_0 + \beta_1 s_{t-1} + \beta_2 s_{t-2} + ... + \beta_5 s_{t-5} + \varepsilon_t$$

another way to build

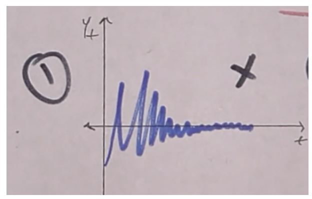
2.2 Stationarity

Definition (Stationarity). The process (X_t) is stationary if

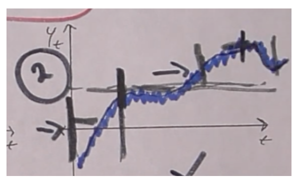
1. $\mathbb{E}(X_t) = \mu$, constant and independent of t
2. $\sigma = cst$
3. No seasonality

Most models like AR or MA assumes the time series is stationary, this is a key concept in time-series.

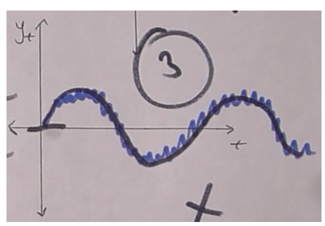
Example. (i) Here the mean is constant, there is no seasonality but the problem is that the std is decreasing over time.



(ii) There is no seasonality but the mean is not constant over time



(iii) The mean and std are constant but there is a seasonality



Difference white nose/stationary. If a time series is white noise we also know it is stationary. The properties are similar σ must be a constant, there is no seasonality but $\mu = 0$ instead of a constant.

Check for stationary. (a) graphically (b) Global vs local test. Compare the global mean and compare the mean if you look on an interval, you can do the same as std and seasonality (c) Augmented dickey-fuller test (ADF test)

tricks to make the series stationary). Suppose we have a time series

$$y_t = \beta_0\beta_1 t + \varepsilon_t \quad \varepsilon \sim \mathcal{N}(0, k^2)$$

This is what we see here

It is not stationary the mean is moving up. The solution is to create a new series the difference between consecutive value of X_t (you can think of it as the difference in height)

$$Z_t = X_t - y_{t-1}$$

We replace

$$
\begin{aligned}
&= (\beta_0 + \beta_1 t + \varepsilon_t) - (\beta_0 + \beta_1(t-1) + \varepsilon_{t-1}) \\
&= \beta_1 + (\varepsilon_t - \varepsilon_{t-1})
\end{aligned}
$$

We can take the expected value

$$\mathbb{E}[Z_t] = \beta_1$$

The mean of this new series is constant.

$$Var(Z_t) = 2k^2$$

Because the errors are uncorrelated some point to the next we just sum them.

What we do is the operation in Z_t then we go back to X_t.

2.3 White Noise: An example of a stationary process

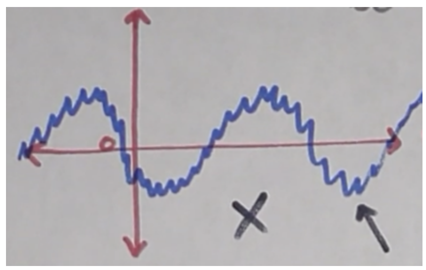
It answers the question of when should I stop fitting my model. You have a residual from your best model and you could ask if you could do better if there is a pattern on the residuals that I should have captured or I am truly done ?

Definition (White Noise). A process (ε_t) will be termed as a White noise (weak) if it is stationary, centered and without autocorrelation

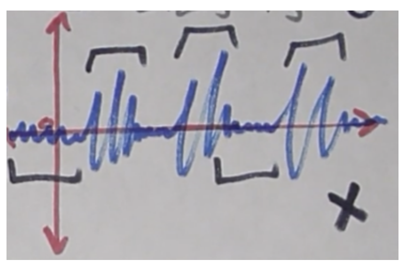
$$\begin{aligned}&\mathbb{E}\left(\varepsilon_t\right) = 0, \forall t\\&V\left(\varepsilon_t\right) = \mathbb{E}[\varepsilon_t^2] = \sigma^2 \text{ and}\\&\rho_\varepsilon(h) = cov(\varepsilon_t, \varepsilon_{t-h}) = \mathbb{E}(\varepsilon_r \varepsilon_{t-h)} = 0 \text{ for } h \neq 0\end{aligned}$$

Aka a time series is a white noise if the mean is 0, the std is constant and the correlation betweeen lags is 0, it means there is no correlation and a lag version of the time series (think of autocorrelation).

Examples (i) mean is zero, std is constant, the correlation is violated because if you are a specific time step it is highly likely that the neighbor will be around a similar value



(ii) std is not constant over time, there is period of high volatility and other of low volatlity. For correlation it is has to say.



Next. A big property of white noise is that "it is not predictable"; Important cause when we do a time series problem we always have

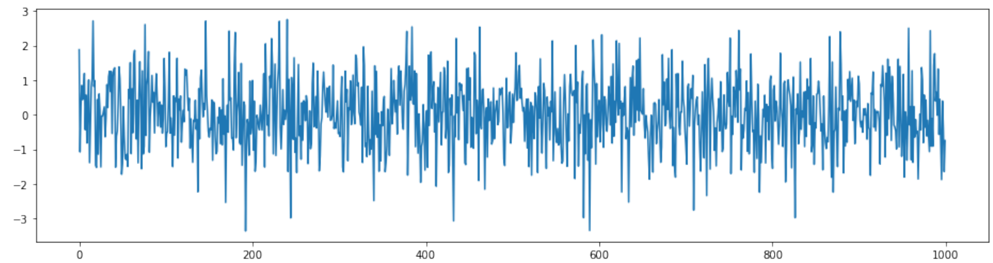
$$X_t = \text{signal} + \text{noise}$$

If you capture a time series perfectly then when you compute the residuals

$$X_t - \text{signal} = \text{noise}$$

The "noise" or residual should be a white noise that is unpredictable.

Test for white noise. (1) visual test (2) global vs local checkers you compare the mean on the whole data and then a rolling window for each window you compute mean and it should match the pop, (cause mean=0, std = cst) (3) check ACF: (correlation lags should be 0)



For a white noise the lag is directly around zero.