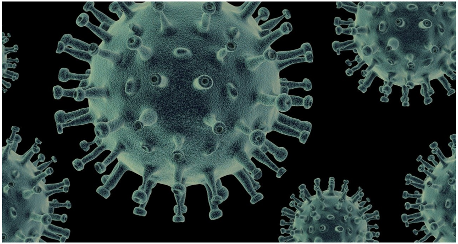
Partial Autocorrelation Function (PACF).

November 10, 2020



Stationary time series type AR(p)

We consider the time series AR(p) in the form:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \epsilon_t,$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are unknown parameters. Simple example shows that AR(p) may be stationary or not:

- the model $AR(1)$ in the form $X_t = X_{t-1} + \epsilon_t$ is not stationary;
- the model $AR(1)$ in the form $X_t = \phi X_{t-1} + \epsilon_t$ is stationary if and only if $\phi \in (-1, 1)$;

Stationary time series type AR(p)

- Assume that our model AR(p) is stationary.
- We find the preliminary estimator of $\phi_1, \phi_2, \ldots, \phi_p$ using Yule-Walker formula.

Yule-Walker formula

We have derived the **Yule-Walker formula**:

$$\underbrace{\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_p \end{bmatrix}}_{\tilde{\rho}} = \underbrace{\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix}}_{\mathbf{R}_\rho} \underbrace{\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_p \end{bmatrix}}_{\phi}$$

Remarks:

1. The matrix $\mathbf{R}_\rho$ - of correlations of the vector $(X_1, X_2, \dots, X_n)$ is positively definite, hence nonsingular;
2. Hence there exists an inverse matrix $\mathbf{R}_\rho^{-1}$;
3. The vector ϕ is uniquely determined $\phi = \mathbf{R}_\rho^{-1}\tilde{\rho}$, i.e. only such a vector determines a stationary model type AR(p);

The preliminary estimation of parameters of stationary time series AR(p)

Since we observe $X_1, X_2, \ldots, X_n$ we can obtain the sampling autocorrelations:

$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0},$$

where for $h \geq 0$ we have

$$\hat{\gamma}_h = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X})$$

and

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

The preliminary estimation of parameters of stationary time series AR(p)

Substituting ρ_h by its estimator $\hat{\rho}_h$ the Yule-Walker furmula provide

$$\underbrace{\begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\rho}_3 \\ \vdots \\ \hat{\rho}_p \end{bmatrix}}_{\hat{\tilde{\rho}}} = \underbrace{\begin{bmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-2} \\ \hat{\rho}_2 & \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \cdots & 1 \end{bmatrix}}_{\mathbf{R}_{\hat{\rho}}} \underbrace{\begin{bmatrix} \hat{\phi}_1^{YW} \\ \hat{\phi}_2^{YW} \\ \hat{\phi}_3^{YW} \\ \vdots \\ \hat{\phi}_p^{YW} \end{bmatrix}}_{\hat{\phi}^{YW}}$$

Multiplying both sides by the corresponding inverse matrix we have the **preliminary estimator** of $\phi_1, \phi_2, \ldots, \phi_p$,

$$\hat{\phi}^{YW} = \mathbf{R}_{\hat{\rho}}^{-1} * \hat{\tilde{\rho}}.$$

The preliminary estimation of parameters of stationary time series AR(p)

The $\hat{\phi}^{YW}$ is a preliminary estimator of ϕ and cannot be a final estimation due to the following reasons:

1. The model AR(p) is not always stationary, in practice the assumption of stationarity is not reasonal;
2. Even if the model AR(p) is stationary the value of $\hat{\phi}^{YW}$ is sensitive to rounding errors, especially if the parameters of AR(p) are close to the *edge* of stationarity;
3. But because we usually are not aware the value of p, $\hat{\phi}^{YW}$ is usefull as a preliminary estimator, which we apply for determining p (the details later);
4. The final estimator of ϕ will be determined using the *maximum likelihood method* method.

Yule Walker formula. What else?

If X_t type AR(p) in the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \epsilon_t$$

can be embedded into AR(p+1) as well:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + 0 * X_{t-(p+1)} + \epsilon_t$$

and consequently in the arbitrary $AR(p+m)$

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} \\ &+ \underbrace{\phi_{p+1} X_{t-(p+1)} + \phi_{p+2} X_{t-(p+2)} + \ldots + \phi_{p+m} X_{t-(p+m)}}_{\text{Here } \phi_{p+1}=\phi_{p+2}=\ldots=\phi_{p+m}=0} \\ &+ \epsilon_t. \end{aligned}$$

Yule Walker formula. Intuition.

- Now we can say X_t is AR(p) if

$$\phi_p \neq 0 \quad \text{and} \quad \phi_{p+1} = \phi_{p+2} = \ldots = 0,$$

  hence ϕ_p is the last non-zero coefficient;
- If X_t is really type $AR(p)$, but we compute Yule-Walker formula AR(p+m) for some $m > 0$ one can obtain:
  - $\hat{\phi}_p^{YW}$ relatively far from 0, and all next coefficients $\hat{\phi}_j^{YW}$ ($j > p$) should be relatively close to 0:
  - intuitively, whe should obtain

$$|\hat{\phi}_p^{YW}| > \text{a threshold }, \hat{\phi}_{p+1}^{YW} \approx 0, \hat{\phi}_{p+2}^{YW} \approx 0, \ldots, \text{ and } \hat{\phi}_{p+m}^{YW} \approx 0,$$

    and the values $\hat{\phi}_1^{YW}, \hat{\phi}_2^{YW}, \ldots, \hat{\phi}_{p-1}^{YW}$ are less informative for our analysis.

Yule Walker formula. The meaning and the road-map.

For using Yule-Walker formula:

1. We presume X_t is stationary and is in type of $AR(p)$, but we are not aware the size of p;
2. We are able to predict the real value of p;
3. Having the prediction p, we pass to further analysis with $AR(p)$, and we further ignore the value $\hat{\phi}^{YW}$ (it is a preliminary estimator only);
4. We find estimators of $\phi_1, \phi_2, \ldots, \phi_p$ using maximum likelihood method (it will be later).

We construct a **Partial Autocorrelation Function (PACF)**

- We base on time series $X_1, X_2, \ldots, X_n$;
- We apply algorithm of Yule-Walker $n-1$ times successively extending the model from AR(1) through AR(2),AR(3),...,AR(p).
- To distinguish Yule-Walker estimators obtained by distinct models AR(p) we denote

$$\hat{\phi}_{1,p}^{YW} \approx \phi_1, \hat{\phi}_{2,p}^{YW} \approx \phi_2, \ldots \quad , \text{and} \quad \hat{\phi}_{p,p}^{YW} \approx \phi_p.$$

Partial Autocorrelation Function.

Partial Autocorrelation Function (PACF) - algorithm

1. If $p = 1$, the Yule-Walker formula for AR(1) provides
$$\hat{\phi}_{1,1}^{YW} = \hat{\rho}_1 \approx \phi_1;$$
2. If $p = 2$, the Yule-Walker formula for AR(2) provides
$$\hat{\phi}_{1,2}^{YW} \approx \phi_1 \quad \text{and} \quad \hat{\phi}_{2,2}^{YW} \approx \phi_2;$$
3. continue this procedure for any p satisfying $1 \le p \le n - 1$:
$$\hat{\phi}_{1,p}^{YW} \approx \phi_1, \hat{\phi}_{2,p}^{YW} \approx \phi_2, \quad , \ldots, \text{and} \quad \hat{\phi}_{p,p}^{YW} \approx \phi_p.$$
4. The function $p \mapsto \hat{\phi}_{p,p}^{YW}$ is called the **Partial Autocorrelation Function**.

Partial Autocorrelation Function

PACF - application

Assuming that the time series X_t is type AR(p) we compute $\hat{\phi}^{YW}_{p,p}$:

- in next units of lectures we put significance thresholds for $\hat{\phi}^{YW}_{p,p}$;
- if the moduli of $\hat{\phi}^{YW}_{p,p}$ extends this threshold, we say X_{t-p} is insignificant in the computing X_t;
- to accept that the truth model is p one must have
    - the value $\hat{\phi}^{YW}_{p,p}$ is significant;
    - for of $k > p$ most of the values of $\hat{\phi}^{YW}_{k,k}$ are insignificant, and we allow sparse elements of $\hat{\phi}^{YW}_{k,k}$ to be significant provided that they are close to significance threshold.

COVID in Poland 1.10.2020-9.11.2020

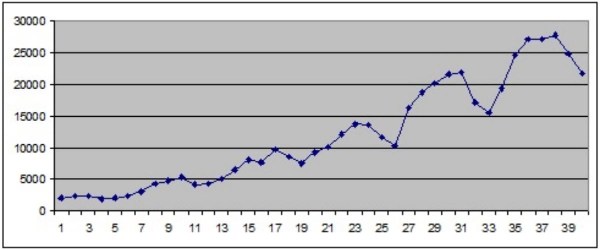
Unfortunately there is no country free of corona-virus COVID-19. The daily number of reported infections in Poland from (1.10.2020) until yesterday (9.11.2020) is reported in the table in the next page. Let X_t be the time series which means the number of infected people in day t, where t the natural index of days in considered period.

COVID in Poland during the fall 2020

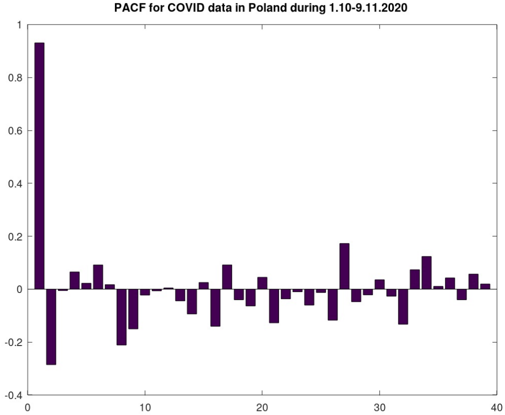
The number of reported infections in Poland during 1.10-9.11.2020.

Day	1-4.10	5-11.10	12-18.10	19-25.10	26.10-1.11	2-8.11	9.11
Monday		2006	4324	7482	10241	15578	21713
Tuesday		2236	5068	9291	16300	19364	
Wednesday		3003	6526	10040	18820	24692	
Thursday	1967	4280	8099	12107	20156	27143	
Friday	2292	4739	7705	13632	21629	27086	
Saturday	2367	5300	9622	13628	21897	27876	
Sunday	1934	4178	8536	11742	17171	24785	

The plot of the trajectory of X_t in the following picture:



Analysis in Octave yields us the following bar plot of PACF for the time series X_t. The height of the bar at p means the value $\hat{\phi}^{YW}_{p,p}$.



The dominating value $\hat{\phi}^{YW}_{1,1}$ over all leads us to suppose that the model is AR(1).

Preliminary corollaries

- Based on analysis of PACF only we may presume X_t has AR(1) form;
- Yule-Walker formula yields $\phi_{1,1}^{YW} = 0.9309$ hence the model is

$$X_t = 0.9309 * X_{t-1} + \epsilon_t;$$

- But Yule-Walker formula always assume stationarity of X_t which is not always acceptable;
- This model is not final and will be corrected using the maximal likelihood method, where we relax the assumption of stationarity and include a constant value:

$$X_t = \mu + \phi X_{t-1} + \epsilon_t;$$

- With the calculus we have

$$X_t = 11866.6 + 0.972640 * X_{t-1} + \epsilon_t;$$

Preliminary corollaries

Remark

Both models should be diagnostically verified and at this stage this model is far to be believable. For example

- we have not compared PACF with the thresholds of significance;
- we have not verified the normality of residua, equality of variance, missing of correlation;
- we have not taken into consideration other phenomena like seasonality (e.g. on Monday we have always less infected people than on Wednesday);
- **in fact we do not know whether dependence of infected today and tomorrow have linear form.**

Because of that, please do not believe this model. It is very preliminary.

But if the following model

$$X_t = 11866.6 + 0.972640 * X_{t-1} + \epsilon_t;$$

is correct, then the expectation is

$$EX_t = 11866.6 + 0.972640 * EX_{t-1}.$$

Then,

- yesterday we had 21 713 infected people;
- tomorrow we expect

$$EX_{41} = 11866.6 + 0.972640 * 21713 = 32985.53;$$

- and the expectation will increase until it reaches per day

$$\lim_{t\to\infty} E_t = \frac{11866.6}{1 - 0.972640} \approx 433\,720$$

- According to this model, there is no evidence that the epidemic is going to slow down, quite the contrary.