DS-GA 3001-001:
Probabilistic Time Series Analysis, Fall 2019
Homework Assignment 1
Yves Greatti - yg390

1. Consider the sample mean of a stationary time series $x_t$, defined as:

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

Compute the variance of this estimate $\text{Var}(\hat{\mu})$ ,as a function of T , and the autocovariance function $\gamma(h)$. By definition for a stationary time series, $\gamma(h) = \text{Cov}(x_{t+h}, x_t)$ and by linearity of the covariance:

$$\text{Var}(\hat{\mu}) = \text{Var}(\frac{1}{T} \sum_t x_t) = \frac{1}{T^2} \text{Var}(\sum_t x_t, \sum_s x_s)$$

$$= \frac{1}{T^2} \big( T\gamma_x(0) + (T-1)\gamma_x(1) + (T-2)\gamma_x(2) + \cdots + \gamma_x(T-1) +$$

$$(T-1)\gamma_x(-1) + (T-2)\gamma_x(-2) + \cdots + \gamma_x(1-T) \big)$$

$$= \frac{1}{T} \sum_{h=-T}^{T} \big(1 - \frac{|h|}{T}\big) \gamma_x(h)$$

2. Confidence bounds for the autocorrelation function: show that the variance of the empirical ACF for white noise with variance $\sigma^2$ given T data points is $\frac{1}{T}$. By Theorem A.7 from ts4.pdf, asymptotic variance of the empirical ACF of stationary process, for T large, is given by:

$$\begin{pmatrix} \begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(h) \end{pmatrix} \end{pmatrix} \approx \text{AN} \begin{pmatrix} \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(h) \end{pmatrix}, T^{-1}\boldsymbol{W} \end{pmatrix}$$

where:

$$W_{pq} = \sum_{u=1}^{\infty} [\rho(u+p) + \rho(u-p) - 2\rho(p)\rho(u)] \times [\rho(u+q) + \rho(u-q) - 2\rho(q)\rho(u)]$$

Given a white noise with variance $\sigma^2$, $\rho(0) = 1$, $\rho(h) = 0$ for $|h| > 0$ , then in the sum for $W_{pq}$, the only surviving term is $u = p = q$, i.e. $W_{pp} = 1$, $W_{pq} = 0$ for $p \neq q$. Hence

$$\begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(h) \end{pmatrix} \approx \mathcal{N}(0, T^{-1}I)$$

and $\hat{\rho}(.) \approx \mathcal{N}(0, T^{-1})$.

3. For an MA(1), $x_t = w_t + \theta w_{t-1}$ show that the autocorrelation function $|\rho_x(1)| \leq 0.5$, for all $\theta$. For which values $\theta$ is it maximum/minimum?

*Solution:* The ACF is defined as:

$$\rho_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)}$$

For h =1

$$\begin{aligned} \gamma_x(t+1, t) &= \mathrm{Cov}(w_t + \theta w_{t-1}, w_{t+1} + \theta w_t) \\ &= \theta \cdot \mathrm{Cov}(w_t, w_t) \\ &= \theta \sigma^2 \end{aligned}$$

Since $w_t$ is a white noise with variance $\sigma$ and $\mathrm{Cov}(w_s, w_t) = 0$ for $s \neq t$. And for h=0

$$\begin{aligned} \gamma_x(t, t) &= \mathrm{Cov}(w_t + \theta w_{t-1}, w_t + \theta w_{t-1}) \\ &= \mathrm{Cov}(w_t, w_t) + \mathrm{Cov}(\theta w_{t-1}, \theta w_{t-1}) \\ &= \sigma^2 + \theta^2 \sigma^2 \\ &= (1 + \theta^2) \cdot \sigma^2 \end{aligned}$$

Thus $\rho_x(1) = \frac{\theta}{(1+\theta^2)}$, $(\theta - 1)^2 \geq 0 \Rightarrow \theta^2 - 2\theta + 1 \geq 0 \Rightarrow |\frac{\theta}{1+\theta^2}| \leq \frac{1}{2} \Rightarrow |\rho_x(1)| \leq 0.5$. Let $f(x) = \frac{x}{1+x^2}$ then $f'(x) = \frac{(1-x)(1+x)}{(1+x^2)^2}$ and $f'(x) = 0$ for $x = 1$ or $x = -1$. $f''(x) = 2\frac{x(x^2-3)}{(1+x^2)^3}$ and $f''(1) = -\frac{1}{2} < 0$ and $f''(-1) = 1 > 0$. $f(1) = f(-1) = \frac{1}{2}$ and $\theta = 1$ and $\theta = -1$ are the global maximum and minimum.

4. Identify the following models as ARMA(p,q):

- $x_t = 0.8x_{t-1} - 0.15x_{t-2} + w_t - 0.3w_{t-1}$
- $x_t = x_{t-1} - 0.5x_{t-2} + w_t - w_{t-1}$

(a) Using backshift operator $\mathbf{B}$, we can rewrite the first model as: $(1 - 0.8\,\mathbf{B} + 0.15\mathbf{B}^2)x_t = (1 - 0.3\,\mathbf{B})w_t$ The roots of the polynomial $\Theta(x) = 1 - 0.8x + 0.15x^2$ are $\theta_1 = 2$ and $\theta_2 = \frac{1}{0.3}$. $\Rightarrow \Theta(x) = \frac{1}{0.3}(\mathbf{B} - 2)(0.3\,\mathbf{B} - 1)$. After simplification this is a ARMA(1,0) or AR(1) model.

(b) Similarly we can rewrite the second model as: $(1 - \mathbf{B} + 0.5\mathbf{B}^2)x_t = (1 - \mathbf{B})w_t$, setting $\Theta(x) = 1 - x + 0.5x^2$ we have:

$$\begin{aligned} \Theta(x) &= \frac{1}{2}(x^2 - 2x + 2) \\ &= \frac{1}{2}(x - x_1)(x - x_2) \text{ where } x_1 = 1 - i \text{ and } x_2 = 1 + i \end{aligned}$$

There are no simplifications on the left or right sides of the model and no parameters are redundant therefore this model is an ARMA(2,1).

2

5. Having observed a sequence $\{x_1, x_2...x_t\}$ we are trying to predict a future observation $x_{t+h}$, with $h \geq 1$. How well / far can one predict into the future if the data comes from a a) MA(3) and b) AR(1) model.

   *Hint: think of the functional form of the optimal estimator and/or the corresponding graphical model.*

   We have seen in class that forecasting consists in determining the parameters of the model, and from there the autocovariance function for different lags, and then the autocorrelation function, to end with the covariance matrix of the model. With the covariance matrix known, we can use it to make inference (forecasting).

   The autocovariance function of a MA(q) process is $\gamma(h) = \sigma^2 \sum_{i=0}^{q-h} \theta_i \theta_{i+h}$ for $h = 1, \cdots, q$ and $\gamma(h) = 0$ for $h > q$ where $\theta_0 = 1$. That means that a future prediction $x_\tau$ is correlated at most by the last three observed samples $\{x_{\tau-3}, x_{\tau-2}, x_{\tau-1}\}$. This result can be observed when plotting the ACF of the MA(3) process, on which you can see that at most 3 spikes are non-zero. With an optimal estimator of a MA(3) process, we can forecast up to 3 out-of sample periods.
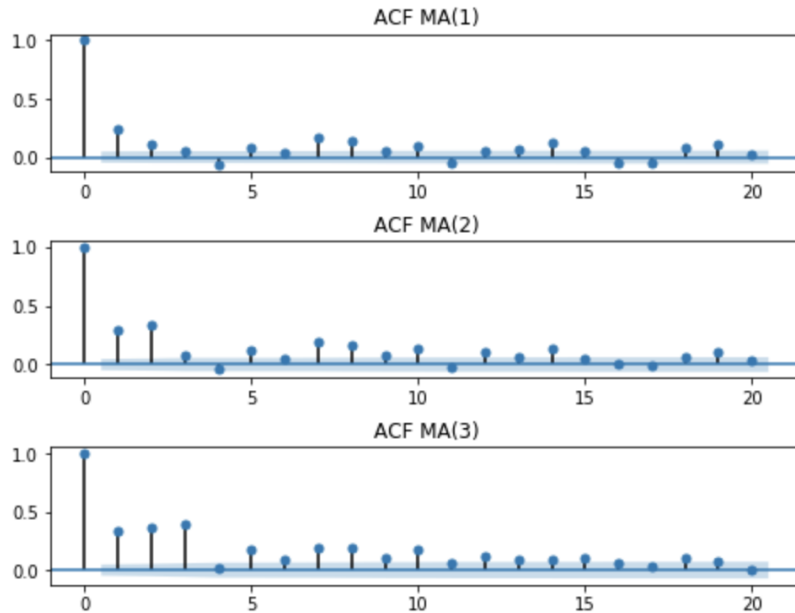


   In the case of a AR(p) such as a AR(1) process $x_t = \theta x_{t-1} + w_t$, the PACF identifies the order of the process. The corresponding graphical model shows that $x_t$, is conditionally depending on $x_{t-1}$ and $w_t$. A numerical plots of empirical PACF of three AR(1), AR(2) and AR(3) processes, confirms that only one spike is different than zero for a AR(1), two pikes for a AR(2) and three spikes for a AR(3) process. So for a AR(1) process we can forecast up to one-step ahead.
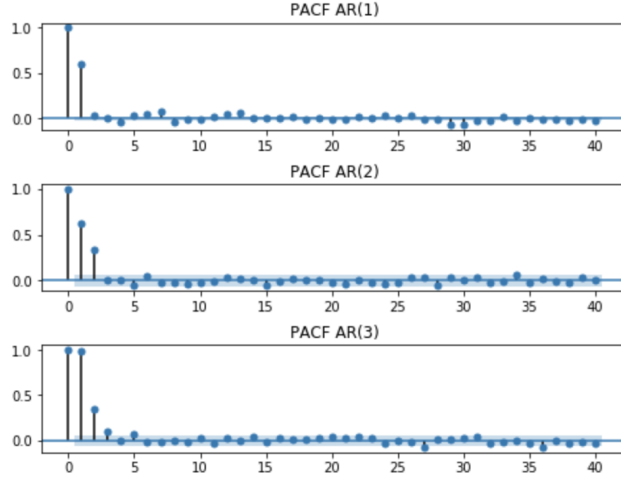
6. Given the AR(2) process with $P(\mathbf{B}) = (1 - 0.2\,\mathbf{B})(1 - 0.5\,\mathbf{B})$, what is $\rho(h)$? Check your analytical solution against an empirical estimate using the code from the lab. The ACF of an AR(2) process is defined as:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$$

$$\mathrm{E}\left(x_t x_{t-h}\right) = \phi_1\,\mathrm{E}\left(x_{t-1} x_{t-h}\right) + \phi_2\,\mathrm{E}\left(x_{t-2} x_{t-h}\right) + \mathrm{E}\left(w_t x_{t-h}\right)$$

$$\mathrm{E}\left(w_t x_{t-h}\right) = \mathrm{E}\left(w_t \sum_{j=0}^{\infty} \psi_j w_{t-h-j}\right) = 0 \Rightarrow$$

$$\gamma(h) = \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2) \text{ for } h = 1, 2, \cdots \Rightarrow$$

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2) \text{ by dividing by } \gamma(0)$$

where we have used $\mathrm{E}\left(w_t w_s\right) = 0$ for $t \neq s$. In the given example, the AR(2) process is $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$ where $\phi_1 = 0.7$ and $\phi_2 = -0.1$. When checking the analytical solution against empirical estimate for $\rho(1)$ and $\rho(2)$, we have $\rho(0) = 1$ and $\rho(h) = \rho(-h)$, applying the previous equality to $\rho(1)$, we obtain $\rho(1) = \phi_1 \rho(0) + \phi_2 \rho(1)$, and $\rho(2) = \phi_1 \rho(1) + \phi_2$:

$$\rho(1) = \frac{\phi_1}{1 - \phi_2}$$

$$\rho_2 = \frac{\phi_1^2 + (1 - \phi_2)\phi_2}{1 - \phi_2}$$
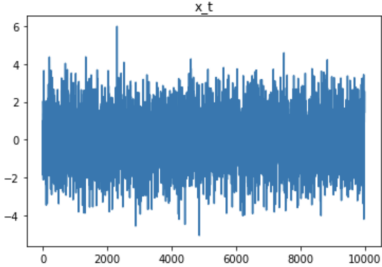
Using the code from the lab:

```
[34]: def acf_impl(x, nlags):
          """
          TODO
          @param x: a 1-d numpy array (data)
          @param nlags: an integer indicating how far back to compute the ACF
          @return a 1-d numpy array with (nlags+1) elements.
                  Where the first element denotes the acf at lag = 0 (1.0 by definition).
          """
          #TODO: replace the template code with your code here. This part will be graded.
          def r(x, y):
              return np.mean((x - x.mean()) * (y - y.mean())) / (x.std() * y.std())

          return np.array([1] + [r(x[:-i], x[i:]) for i in range(1, nlags)])
```

```
[23]: n = 10000
      mean = 0
      std = 1
      lag = 2
      w_t = np.random.normal(mean, std, size=n)
```

```
[44]: x_t = np.zeros(n)
      x_t[0] = 0.5
      x_t[1] = 1
      for i in range (2, n):
          x_t[i] = 0.7 * x_t[i-1] - 0.1 * x_t[i-2] + w_t[i]
```

```
[6]: plt.plot(x_t)
     plt.title("x_t")
     plt.show()
```



```
[7]: acf_val = acf(x=x_t, nlags=lag, fft=True)
     print("StatsModels ACF={}".format(acf_val))
```

StatsModels ACF=[1.        0.6380166 0.3451316]

```
[8]: phi1 = 0.7
     phi2 = -0.1
```

```
[9]: r1 = phi1/(1-phi2)
     r2 = ((phi1 * phi1) + (1-phi2) * phi2)/(1-phi2)
     print(r1, r2)
```

0.6363636363636362 0.34545454545454535

```
[10]: %%latex
      $\rho(1)=$0.6363636363636362 $\rho(2)=$=0.34545454545454535
```

$\rho(1) =$ 0.6363636363636362 $\rho(2) ==$0.34545454545454535