DS-GA 3001.001 Special Topics in Data Science: Modeling Time Series
Homework 1

Due date: February 12, by 5 pm

Problem 1. Consider the periodic time series (period determined by parameter ω) constructed as:

$$x_t = U_1 \sin(2\pi\omega t) + U_2 \cos(2\pi\omega t)\,, \tag{1}$$

with $U_{1,2}$ independent random variables with zero mean and variance σ^2. Show that this series is weakly stationary with autocovariance function $\gamma(h) = \sigma^2 \cos(2\pi\omega h)$. Optional: Is this series strictly stationary?

Solution:

First we need to show that the mean of the series is time independent:

$$\begin{aligned}
\mathbb{E}[x_t] &= \mathbb{E}[U_1 \sin(2\pi\omega t) + U_2 \cos(2\pi\omega t)] \\
&= \mathbb{E}[U_1] \sin(2\pi\omega t) + \mathbb{E}[U_2] \cos(2\pi\omega t) \\
&= 0
\end{aligned}$$

where we have used the fact that the mean of $U_{1,2}$ is zero.

Second, we need to show that the autocovariance function depends only on the time difference h:

$$\begin{aligned}
\mathbb{E}[x_t x_{t+h}] &= \mathbb{E}[(U_1 \sin(2\pi\omega t) + U_2 \cos(2\pi\omega t))\,(U_1 \sin(2\pi\omega(t+h)) + U_2 \cos(2\pi\omega(t+h)))] \\
&= \mathbb{E}[U_1^2] \sin(2\pi\omega t) \sin(2\pi\omega(t+h)) \\
&+ \mathbb{E}[U_1 U_2] \sin(2\pi\omega t) \cos(2\pi\omega(t+h)) \\
&+ \mathbb{E}[U_1 U_2] \sin(2\pi\omega(t+h)) \cos(2\pi\omega t) \\
&+ \mathbb{E}[U_2^2] \cos(2\pi\omega t) \cos(2\pi\omega(t+h)) \\
&= \sigma^2 (\cos(2\pi\omega t) \cos(2\pi\omega(t+h)) + \sin(2\pi\omega t) \sin(2\pi\omega(t+h))) \\
&= \sigma^2 \cos(2\pi\omega h)
\end{aligned}$$

where we have used the fact that $U_{1,2}$ are independent (mixed terms are zero), their second moment is $\mathbb{E}[U_i^2] = \sigma^2 - \mathbb{E}[U_i]^2 = \sigma^2$, and lastly the trigonometric formula $\cos(a-b) = \cos(a)\cos(b) + \sin(a)\sin(b)$.

Note: Since the other moments of the distribution are unspecified, we cannot prove that the series is strictly stationary.

Problem 2. 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 θ. For which values θ is it maximum and minimum?

Solution:

We have already computed the autocovariance for an MA(1) in Lecture 2 (but bonus points if you've included the full derivation):

$$R_X(h) = \begin{cases} \sigma^2(1+\theta^2) & , h = 0 \\ \sigma^2\theta & , |h| = 1 \\ 0, & |h| > 1 \end{cases}$$

After the normalization, the autocorrelation function becomes:

$$\rho_X(h) = \begin{cases} 1 & , h = 0 \\ \frac{\theta}{1+\theta^2} & , |h| = 1 \\ 0, & |h| > 1 \end{cases}$$

To prove that $|\rho_x(1)| \le 0.5$, we multiply both sides by the denominator and rearrange all terms to the right, giving $\theta^2 - 2|\theta| + 1 \ge 0$ which can be rewritten as $(|\theta| - 1)^2 \ge 0$ q.e.d. To find the extrema points, we take the first derivative and set it to zero, which gives: $\frac{1-\theta^2}{(1+\theta^2)^2} = 0$, with roots $\theta_{\text{min/max}} = \pm 1$ and the corresponding optima ± 0.5.

Problem 3. Identify the following models as ARMA(p, q):

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

Note: watch out for parameter redundancy!

Solution:
We first write the expression in canonical form (AR terms left of the = MA terms to the right), $x_t - 0.8x_{t-1} + 0.15x_{t-2} = w_t - 0.3w_{t-1}$, then identify the corresponding P,Q polynomials :

$$\begin{aligned} P(B) &= 1 - 0.8B + 0.15B^2 = (1 - 0.3B)(1 - 0.5B) \\ Q(B) &= 1 - 0.3B \end{aligned}$$

Since the two share a factor, we can simply the definition of the model to

$$\begin{aligned} P(B) &= 1 - 0.5B \\ Q(B) &= 1 \end{aligned}$$

which translates into $x_t = 0.5x_{t-1} + w_t$ which is an AR(1) process. Note: P's root is $2 > 1$, so the process is causal.

Similarly, the second model $x_t - x_{t-1} + 0.5x_{t-2} = w_t - w_{t-1}$ has corresponding polynomials

$$\begin{aligned} P(B) &= 1 - B + 0.5B^2 \\ Q(B) &= 1 - B \end{aligned}$$

These don't have common roots so the model is an ARMA(2,1); P has complex roots $1 \pm i$ which are outside the unit circle, so it's a causal process. Q has root 1 so it's not invertible.

Problem 4. Given an AR(1) model $x_t = \phi x_{t-1} + w_t$, determine the form of m-step ahead prediction $x^t_{t+m} = \mathbb{E}[x_{t+m}|x_t]$ and show that the variance of this estimate has the form:

$$\mathbb{E}[\left(x^t_{t+m} - x_{t+m}\right)^2] = \sigma_w^2 \frac{1 - \phi^{2m}}{1 - \phi^2} \tag{2}$$

Solution:
The joint distribution for (x_t, x_{t+m}) is multivariate normal with mean $[0,\ 0]$ and covariance $\begin{bmatrix} R_x(0) & R_x(m) \\ R_x(m) & R_x(0) \end{bmatrix}$ where the AR(1) autocovariance is $R_x(h) = \sigma^2 \frac{\phi^h}{1-\phi^2}$ (from lecture or rederived).
The posterior distribution $x_{t+m}|x_t$ is also gaussian with mean $\mu_x = \phi^m x_t$ and variance $R_x(0) - R_x(m)^2/R_x(0)$ (using the multivariate gaussian conditioning formula). After replacing in the expression for the covariance this finally gives the variance we were looking for:

$$\mathrm{Var}(x_{t+m}|x_t) = \frac{\sigma^2}{1-\phi^2} - \left(\sigma^2 \frac{\phi^m}{1-\phi^2}\right)^2 \left(\frac{\sigma^2}{1-\phi^2}\right)^{-1} = \frac{\sigma^2}{1-\phi^2}(1 - \phi^{2m}) \tag{3}$$

Problem 5. Consider the datasets recording weekly *oil* (dollars/barrel) and *gas*(cents/gallon) provided in the data folder.

- plot data on same graph; which of the simulated series in lecture 1 do these series resemble? Do you believe they are stationary? (explain why/ why not)

- argue that the transformation $y_t = \nabla x_t$ might be useful for the oil data. Plot y_t, estimate sample ACF and comment on result.
- construct scatterplot of y_t vs. y_{t+h} for the two datasets, $h = \{1, 2, 3\}$. Are these dependencies linear?

Anything sensible is fine here, but it would be nice to see some attempts to check the statistical structure beyond simple visualisation - e.g. fitting a regression model for the integrated part , or looking at the histogram of residuals to check for gaussianity or any other sensible rationalization of the form "if data were coming from an ARMA model then... "