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

Due date: September 27, by 6pm

Problem 1. (5pt) Consider the *sample mean* of a stationary time series x_t, defined as:

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

Compute the variance of this estimate $Var[\hat{\mu}]$, as a function of T, and the autocovariance function $\gamma(h)$.
Hint: The empirical mean is also a linear combination of random variables, so you can use the formula for the covariance of linear combinations of random variables from the lecture.

Solution:
The empirical mean is a linear combination of (in the ARMA world, gaussian) random variables:

$$\hat{\mu} = \frac{1}{T} \sum_t x_t. \tag{2}$$

so the variance of $\hat{\mu}$ can be expressed (using the usual formula for the covariance of linear combinations of random variables from the lecture) as:

$$\mathrm{Var}[\hat{\mu}] = \mathrm{Cov}\left[\frac{1}{T}\sum_t x_t, \frac{1}{T}\sum_s x_s\right] = \frac{1}{T^2}\sum_{t,s}\mathrm{Cov}[x_t, x_s] = \frac{1}{T^2}\sum_{t,s}\rho(t-s) \tag{3}$$

Lots of terms in that sum, but we can group them by lag, h, and then count the number of (t, s) pairs with the same h, which yields:

$$\mathrm{Var}[\hat{\mu}] = \frac{1}{T^2}\sum_{h=-T}^{T}(T - |h|)\rho(h) \tag{4}$$

As a sanity check, we can see that in the case of iid data this reduces to the usual $1/T$ scaling in CLT.

Problem 2. (10pt) Confidence bounds for the autocorrelation function: show that the variance of the empirical ACF for white noise with variance σ^2 estimated given T data points is $\frac{1}{T}$.
Hint: Use theorem A.7 from tsa4.pdf; alternatively, you can just show it numerically by plotting empirical estimates of the ACF as a function of T.

Solution:
If using A.7: for any stationary linear process, the estimate of the ACF, $\hat{\boldsymbol{\rho}}$ is asymptotically normal, with mean $\boldsymbol{\rho}$ and covariance $\mathbf{W}$, with elements (A.54):

$$W_{pq} = \frac{1}{T}\sum_u \left[\rho(u+p) + \rho(u-p) - \rho(u)\rho(p)\right]\left[\rho(u+q) + \rho(u-q) - \rho(u)\rho(q)\right]. \tag{5}$$

In the special case of white noise, we have $\rho(h) = 1$ for $h = 0$ and zero otherwise; the only way to make one of the factors above nonzero is to have $u = p$ or $u = q$, but the two conditions need to be satisfied simultaneously for the product of the two to be nonzero. Hence for white noise the covariance becomes the rescaled identity:

$$\mathbf{W} = \frac{1}{T}\mathbf{I}$$

. The confidence bounds, typically defined as 2xstd are hence $2/\sqrt{T}$.

The same kind of scaling can be seen numerically by varying T and for each sequence length repeatedly sampling from white noise sequences, computing the ACF for each, then the variance of this estimate pooling across draws.
N.B. While the proof is asymptotic, the numerics should convince us that the confidence bounds behave reasonably even when T is not that large.

Problem 3. (10pt) 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/minimum?

Solution:
We have already computed the CCF for an MA(1) in Lecture 2 (but even better if including the full derivation):

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

The corresponding ACF is:

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

To prove that $|\rho_x(1)| \leq 0.5$, we multiply both sides by the denominator and rearrange all terms to the right, giving $\theta^2 - 2|\theta| + 1 \geq 0$ which can be rewritten as $(|\theta| - 1)^2 \geq 0$ 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 4. (5pt+5pt) 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, MA terms 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 need to simply the model to:

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

which translates into $x_t = 0.5x_{t-1} + w_t$; this is an AR(1) process. 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: not invertible.

Problem 5. (5pt) 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.

Solution:
Using the graphical model and/or what we know about the ACF of an MA(q) process, we know that a) MA(3) dependencies only span 3 steps, so future observations beyond this point will be independent; b) for an AR(1) x_{t+h} prediction only depends on x_t, the optimal linear scaling decays exponentially in h (exactly how fast will depend on the parameters).

Problem 6. (10pt) Given the AR(2) process with $P(B) = (1 - 0.2B)(1 - 0.5B)$, what is $\rho(h)$? Check your analytical solution against an empirical estimate obtained using the code from the lab.
Hint: Difference equations + initial conditions.

Solution:
Using the difference equation form of the recursion and observing that the AR polynomial has two roots $z_1 = 5$ and $z_2 = 2$ (making it causal), we know that the functional form for the CCF is: $\rho(h) = c_1 z_1^{-|h|} + c_2 z_2^{-|h|}$, where the two constants need to be determined from initial conditions.[1] How do we do that? One obvious initial condition is that $\rho(0) = 1$; the second constraint takes a bit more work. First, let's write out the explicit AR form:

$$x_t = 0.7x_{t-1} - 0.1x_{t-2} + w_t$$

The ACF recursion equation is then:

$$\rho(h) = 0.7\rho(h-1) - 0.1\rho(h-2)$$

Now we take $h = 1$ and take advantage of the fact that $\rho(1) = \rho(-1)$ to get the second constraint:

$$\rho(1) = \frac{7}{11}$$

To finish up, compute the cs and validate the solution numerically (-5/11 and 16/11).

[1]So dependencies decay exponentially fast.