Homework 3
STAT 5511 (Fall 2023)
Charles R. Doss
Assigned: Fri Oct 11
Due: Mon, Oct 21

The usual formatting rules:

- Your homework (HW) should be formatted to be easily readable by the grader.
- You may use knitr or Sweave in general to produce the code portions of the HW. However, the output from knitr/Sweave that you include should *be only what is necessary to answer the question*, rather than just any automatic output that R produces. (You may thus need to avoid using default R functions if they output too much unnecessary material, and/or should make use of `invisible()` or `capture.output()`.)
  - For example: for output from regression, the main things we would want to see are the estimates for each coefficient (with appropriate labels of course) together with the computed OLS/linear regression standard errors and p-values. If other output is not needed to answer the question, it should be suppressed!
- Code snippets that directly answer the questions can be included in your main homework document; ideally these should be preceded by comments or text at least explaining what question they are answering. Extra code can be placed in an appendix.
- All plots produced in R should have appropriate labels on the axes as well as titles. Any plot should have explanation of what is being plotted given clearly in the accompanying text.
- Plots and figures should be *appropriately sized*, meaning they should not be too large, so that the page length is not too long. (The arguments fig.height and fig.width to knitr chunks can achieve this.)
- **Directions for "by-hand" problems:** In general, credit is given for (correct) shown work, not for final answers; so show **all** work for each problem and explain your answer fully.

Questions:

1. ARMA models: Several ARMA models are written below. You can assume in all cases that $W_t \overset{\text{iid}}{\sim} N(0,1)$. For each of the ARMA models, find the roots of the AR and MA polynomials. Identify any parameter redundancy: find the values of p and q for which each model is ARMA(p,q) and write the model in its correct (non-redundant) form. Determine whether each model is causal, and determine whether it is invertible.
   (a) $X_t + 0.81X_{t-2} = W_t + \frac{1}{3}W_{t-1}$
   (b) $X_t - X_{t-1} = W_t - \frac{1}{2}W_{t-1} - \frac{1}{2}W_{t-2}$
   (c) $X_t - 3X_{t-1} = W_t + 2W_{t-1} - 8W_{t-2}$
   (d) $X_t - 2X_{t-1} + 2X_{t-2} = W_t - \frac{8}{9}W_{t-1}$
   (e) $X_t - 4X_{t-2} = W_t - W_{t-1} + \frac{1}{2}W_{t-2}$
   (f) $X_t - \frac{9}{4}X_{t-1} - \frac{9}{4}X_{t-2} = W_t$
2. Linear representation of ARMA: For those models of Question 1 that are causal, compute theoretically the first five coefficients $\psi_0, \ldots, \psi_4$ in the causal linear process representation $X_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}$
3. Autocorrelation function of ARMA model: For the models in parts (a) and (b) of Question 1: Simulate 100 observations from each model. Compute and plot the sample ACF together with the theoretical ACF. You can have R compute the theoretical ACF using the ARMAacf function.
4. (Comparing time series with periodic behavior)
   (a) Find an AR(2) process whose periodic ACF $\rho(h)$ has period 9.
   (b) Simulate a time series following the distribution you found in the previous part. Plot both the *true ACF* and the *simulated data series* (not the sample ACF!).
   (c) Simulate a signal-in-noise time series $Y_t = \mu_t + W_t$ where $W_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$ with $\sigma^2 = 0.01$ (or $\sigma = .1$), and where the signal μ_t is given by the ACF ρ you found in the first part. That is: $\mu_t = \rho(t)$. Plot the *simulated data series*. (If you want you may also plot the underlying true signal $\mu_t = \rho(t)$ on the same plot.)

(d) The point of what we've done so far is to compare different models that lead to periodic behavior. So: discuss/compare both the locations and the frequency/number of the peaks (maxima) and valleys (minima) of the *simulated AR(2) data series* to the locations of the maxima and minima of the *true ACF* and to the locations of the maxima and minima of the *signal in noise data series*.