STA 642: Fall 2018 – Homework #2 Exercises

1. Work through the AR(p) notes and especially the AR(2) examples as introduced in class. Among other things, this provides a foundation for a lot of basic/exploratory analysis of time series with AR models, as well as core theory and more building-blocks for general linear state space models. Read ahead into course notes and slides for the coming week(s), and get intimate with relevant sections of the P&W text.

   In particular, read and digest the material on AR(p) model order assessment using marginal likelihoods and information criteria (AIC, BIC) in P&W section 2.3.4. This relates to the question of selection of, and more broadly "inference on", the AR model order p, with more formal inferential/model-based ideas that complement the exploratory uses of ACF, PACF and other exploratory ideas in regression that you might use.

   The course code repository includes the Matlab function `arpcompare.m`. Open and review that function– it simply codes up the formal Bayesian marginal likelihood computation for model order p based on the conditional reference analysis, and also generates the AIC & BIC measures. It is used in examples in the class examples code for AR(p) models.

2. This excercise adds some theoretical structure to the stationary AR(1)+noise model (a.k.a. hidden Markov model). As part of this, the example here gets into ARMA models, and you may find the use of the backshift operator strategy useful (– though by no means necessary–) to explore parts of this question.

   In the stationary AR(1)+noise model– a first state space/hidden Markov model– we observe

   $$y_t = x_t + \nu_t \quad \text{where} \quad x_t \leftarrow AR(1|(\phi, v))$$

   with $\nu_t \sim N(0, w)$ and assuming $\nu_t \perp\!\!\!\perp \nu_r$ and $\nu_t \perp\!\!\!\perp \epsilon_r$ for all t, r. Clearly $y_t \sim N(0, q)$ with $q = s + w$ where $s = v/(1-\phi^2)$.

   This exercise shows that y_t is *not* an AR(1) process; it is an ARMA(1,1) process as in the resulting AR-like form of the implied model for y_t, the driving innovations are correlated at lag-1.

   (a) Show that $y_t = \phi y_{t-1} + \eta_t$ where $\eta_t = \epsilon_t + \nu_t - \phi\nu_{t-1}$.

   This is most easily seen by applying the backshift operator $\phi(B) = 1 - \phi B$ to both sides of the identity $y_t = x_t + \nu_t$ to give $\phi(B)y_t = \phi(B)x_t + \phi(B)\nu_t$ or just $y_t - \phi y_{t-1} = x_t - \phi x_{t-1} + \nu_t - \phi\nu_{t-1}$ and the result follows since $x_t - \phi x_{t-1} = \epsilon_t$.

   The use of the backshift operator here is, of course, just a shorthand for working the algebra directly. That is– in longhand, by substitution– we have

   $$y_t = x_t + \nu_t = \phi x_{t-1} + \epsilon_t + \nu_t = \phi(y_{t-1} - \nu_{t-1}) + \epsilon_t + \nu_t$$

   and the result follows.

(b) Show that the lag-1 correlation in the η_t sequence is $-\phi w/(w(1+\phi^2)+v)$.

First, $V(\eta_t) = V(\epsilon_t + \nu_t - \phi\nu_{t-1}) = V(\epsilon_t) + V(\nu_t) + \phi^2 V(\nu_{t-1})$ by independence. This reduces to $V(\eta_t) = v + (1+\phi^2)w$. Second,

$$\begin{aligned} Cov(\eta_t, \eta_{t-1}) &= Cov(\epsilon_t + \nu_t - \phi\nu_{t-1}, \epsilon_{t-1} + \nu_{t-1} - \phi\nu_{t-2}) \\ &= Cov(-\phi\nu_{t-1}, \nu_{t-1}), \quad \text{by independence,} \\ &= -\phi V(\nu_{t-1}) \; = \; -\phi w. \end{aligned}$$

The stated correlation follows.

Additional comment: $Cor(\eta_t, \eta_{t-1})$ has the opposite sign to ϕ and is smaller in absolute value.

(c) Find an expression for the lag$-k$ autocorrelation of the y_t process in terms of k, ϕ and the signal:noise ratio s/q. Comment on this result. (We already worked through this in class; do it again!)

The covariance at lag$-k$ is

$$\begin{aligned} E(y_t y_{t-k}) &= E((x_t + \nu_t)(x_{t-k} + \nu_{t-k})) \\ &= E(x_t x_{t-k}) + E(x_t \nu_{t-k}) + E(x_{t-k}\nu_t) + E(\nu_t \nu_{t-k}) \\ &= \phi^k s + 0 + 0 + 0 \end{aligned}$$

so the correlation is $\phi^k s/q$. That is, relative to the lag$-k$ autocorrelation ϕ^k of the underlying signal process, that of the observed data process is directly reduced by the signal:noise ratio $s/q \equiv s/(s+w)$.

(d) Is y_t an AR(1) process? Is it Markov? Discuss and provide theoretical rationalisation.

Credit for any sensible comments or theory, including exploration of graphical model structure as detailed below. One obvious point is that the acf of y_t is *not* that of an AR(1) process (unless $w = 0$)– this shows that y_t is not AR(1), as already discussed in class.

To explore further, see the graphical modelling theory and development below in an appended Question 2. To add to that look in more detail at the structure of the innovations series η_t impacting the y_t process. From above, $\eta_t = \epsilon_t + \nu_t - \phi\nu_{t-1}$ and we already found that $Cov(\eta_t, \eta_{t-1})$ is non-zero. Now, for $k > 2$, the innovation term $\eta_{t-k} = \epsilon_{t-k} + \nu_{t-k} - \phi\nu_{t-k-1}$ has no terms in common with η_t, hence $Cov(\eta_t, \eta_{t-k}) = 0$ for $k > 1$. This means η_t must be a moving-average process of order 1, i.e., MA(1)– and any such process can be written as $\eta_t = \delta_t + \theta\delta_{t-1}$ for some i.i.d., zero mean normal series δ_t, where $|\theta| < 1$ and the lag-1 correlation satisfies $Cor(\eta_t, \eta_{t-1}) = \theta/(1+\theta^2)$. Here θ is known at the moving average coefficient (see Example 2.5, p 65 in P&W). Now, whatever the value of θ may be, this implies that $y_t = \phi y_{t-1} + \delta_t + \theta\delta_{t-1}$, which is an ARMA(1,1) process (Section 2.5 in P&W). This is not a Markov process. It is easiest to see this using the backshift operater B notation as follows. We have $(1-\phi B)y_t = (1-\theta B)\delta_t$ so that

$(1-\phi B)y_t = (1-\theta B)\delta_t$ or $(1-\theta B)^{-1}(1-\phi B)y_t = \delta_t$. Expanding the inverse operator term here gives

$$(1+\theta B+\theta^2B^2+\theta^3B^3+\cdots)(1-\phi B)y_t = \delta_t$$

or

$$\{1+(\theta-\phi)B+\theta(\theta-\phi)B^2+\theta^2(\theta-\phi)B^3+\cdots\}y_t = \delta_t$$

so that

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} \cdots + \delta_t$$

where $\alpha_j = \theta^{j-1}(\theta-\phi)$ for $j = 1, 2, \ldots$. This shows that y_t depends on the full past history of the series, i.e., is an infinite order AR representation.

Here is an extension of discussion of Question 2 to explore graphical model structure. The following is not expected or necessary as part of the solution to the exercise, but is really key as part of the bigger picture. And, by the way, could be presented as a solution to part of the question. The process y_t is not Markovian. While it may seem initially likely– since we are just adding noise to an AR(1) process– the lag-1 dependencies induced in the η_t innovations of the y_t process should raise a doubt. Intuitively, note the following. First, we have shown that η_t is correlated with η_{t-1}; second, η_{t-1} is directly dependent on ν_{t-2} which itself directly influences y_{t-2}; so this suggests y_t is related to y_{t-2} via a path $y_t \leftarrow \eta_t \leftarrow \eta_{t-1} \leftarrow y_{t-2}$ as well as directly via the AR(1)-induced path from $y_{t-1} \leftarrow y_{t-1} \leftarrow y_{t-2}$ (unless $w = 0$, of course). Let us exhibit and explore graphical models for this context.

(a) We know that the (undirected) graph of the full joint distribution $p(y_{1:t}, x_{0:t})$ has the form over $t-4:t$, of, simply,

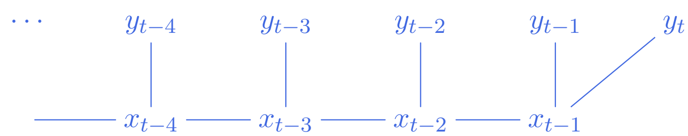
$$\begin{tikzcd}
\cdots & y_{t-4} \arrow[d, dash] & y_{t-3} \arrow[d, dash] & y_{t-2} \arrow[d, dash] & y_{t-1} \arrow[d, dash] & y_t \arrow[d, dash] \\
{} \arrow[r, dash] & x_{t-4} \arrow[r, dash] & x_{t-3} \arrow[r, dash] & x_{t-2} \arrow[r, dash] & x_{t-1} \arrow[r, dash] & x_t
\end{tikzcd}$$

As we know, we "read-off" conditional dependencies based on edges in the graph. For example, we know/see that the conditional distribution for x_t given *all the other quantities up to and including time* t, depends only on its neighbours in the graph, namely x_{t-1}, y_t. Explicitly, x_t is conditionally independent of $x_{0:t-2}, y_{1:t-1}$ given x_{t-1}, y_t.

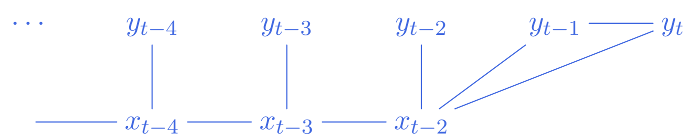
(b) We also know that *marginalization over a variable* just (i) deletes that variable from the graph, and (ii) adds edges between all variables/nodes having an edge to the variable removed. This second point is key: *when I leave, all my neighbours become neighbours (if they are not already)*: this is the graphical recognition that two nodes that may be conditionally independent in the full joint distribution become dependent under marginalization of a variable that "separates" them.

(c) Let us look at marginalization over x_t, to reduce to the marginal distribution $p(y_{1:t}, x_{0:t-1})$. The "join up all my neighbours when I leave" rule results in a new edge between x_{t-1}

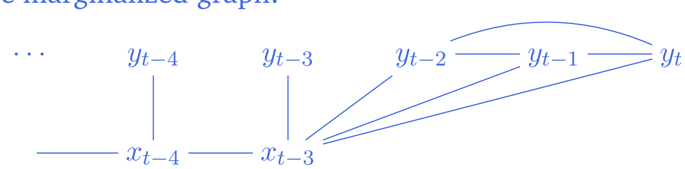
and y_t in the marginalized graph:



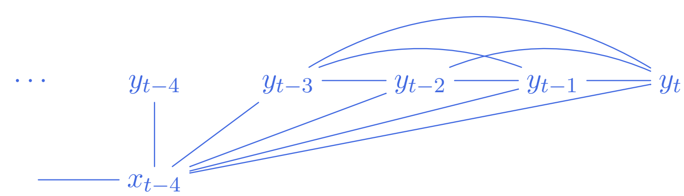
(d) Now marginalize over x_{t-1} to generate the resulting graph of the marginalized distribution $p(y_{1:t}, x_{0:t-2})$. Since x_{t-1} has neighbours x_{t-2}, y_{t-1}, y_t, these nodes all become neighbours in the marginalized graph:



(e) Next marginalize over x_{t-2} to generate the resulting graph of the marginalized distribution $p(y_{1:t}, x_{0:t-3})$. Since x_{t-2} has neighbours $x_{t-3}, y_{t-2}, y_{t-1}, y_t$, they all become neighbours in the marginalized graph:



(f) Take the next step: marginalize over x_{t-3} to define the resulting graph of the marginalized distribution $p(y_{1:t}, x_{0:t-4})$. Joining up all the neighbours of x_{t-3} gives:



(g) Imagine continuing the marginalization process to remove x_{t-4}, then sequentially back in time to remove all x_t down to and including x_0. This will yield the graph of $p(y_{1:t})$, the marginal distribution of the sequence of $y_{1:t}$ marginalized over the hidden signal $x_{0:t}$ altogether. In this distribution, we can then see which past does y_t depend on. That is, in $p(y_t|y_{1:t-1})$ which of the past y_s, $s = 1 : t-1$, matter. Do this by continuing the process above, to see that this induces edges between all pairs of y nodes, so that the resulting graph is a complete graph. Hence $p(y_t|y_{1:t-1})$ depends on *all past* $y_s, s = 1 : t-1$. So we have shown that adding independent noise to a first-order Markov process destroys the Markovian structure. Intuitively, we can see this by remembering that each y_s provides information on the full sequence of x_t values (recall smoothing when inferring the hidden signals). so marginalizing over the signal process links up all of the y_t. We have now shown this directly via the graphical models.

(h) For a series of length n, the variance matrix $V(x_{1:n}) = \mathbf{\Sigma}_n$ has inverse– the precision matrix of the AR(1) process $x_{1:n}$– given by

$$\mathbf{\Sigma}_n^{-1} = v^{-1} \begin{pmatrix} 1 & -\phi & 0 & \cdots & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & \cdots & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & -\phi & 0 \\ 0 & 0 & 0 & -\phi & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & 0 & -\phi & 1 \end{pmatrix}.$$

We also know that the variance matrix of the $y_{1:n}$ process is $\mathbf{\Sigma}_n + w\mathbf{I}$ where $\mathbf{I}$ is the $n \times n$ identity. The above graphical explorations shows us that there are no zeros in the off-diagonal elements of the corresponding precision matrix $(\mathbf{\Sigma}_n + w\mathbf{I})^{-1}$ This is clear since $p(y_t|y_{1:n\backslash t})$ depends on *all* quantities in $y_{1:n\backslash t}$. A zero only occurs in row t, column s when $y_t \perp\!\!\!\perp y_s \,|\, y_{1:n\backslash(t,s)}$, and in this case there are no such occurrences. Incidentally, this shows how simply adding a constant to the diagonal of a matrix with a sparse inverse destroys that sparsity.

Bonus points for developing some of the graphical model aspects.

3. This question relates to an alternative state space representation of an AR(p) model; this is a special case of the state space representation of ARMA models (P&W, top of page 75; note that the AR(p) is the special case when $q = 0, m = p$ in the notation there).

Work this exercise explicitly in the case of $p = 2$; the linear algebra in this special case is easy and the special case illuminating of the more general AR(p) case. Provide solutions to the case of $p = 2$ for assessment. The general case is just a bit more linear algebra and will be given bonus credit, but is otherwise optional.

In the standard state space representation of the AR(p) model we have state vector $\mathbf{x}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$ and model equations $y_t = \mathbf{F}'\mathbf{x}_t$ and $\mathbf{x}_t = \mathbf{G}\mathbf{x}_{t-1} + \mathbf{F}\epsilon_t$ where

$$\mathbf{F} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{G} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

with AR parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ and innovations $\epsilon_t \sim N(0, v)$.

Define the $p \times p$ symmetric matrix $\mathbf{A}$ by

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \phi_2 & \phi_3 & \phi_4 & \cdots & \phi_{p-1} & \phi_p \\ 0 & \phi_3 & \phi_4 & \phi_5 & \cdots & \phi_p & 0 \\ \vdots & \vdots & \vdots & & \iddots & \iddots & \vdots \\ \vdots & \vdots & \vdots & \iddots & \iddots & & \vdots \\ 0 & \phi_{p-1} & \phi_p & 0 & \cdots & \cdots & 0 \\ 0 & \phi_p & 0 & 0 & \cdots & \cdots & 0 \end{pmatrix}$$

(a) Verify that the matrix product $\mathbf{AG}$ is given by

$$\mathbf{AG} = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \phi_4 & \cdots & \phi_{p-1} & \phi_p \\ \phi_2 & \phi_3 & \phi_4 & \phi_5 & \cdots & \phi_p & 0 \\ \phi_3 & \phi_4 & \phi_5 & \phi_5 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \iddots & \iddots & \vdots \\ \vdots & \vdots & \vdots & \iddots & \iddots & & \vdots \\ \phi_{p-1} & \phi_p & 0 & 0 & \cdots & \cdots & 0 \\ \phi_p & 0 & 0 & 0 & \cdots & \cdots & 0 \end{pmatrix}$$

noting that this is also symmetric.

Special case of $p = 2$: In this case

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & \phi_2 \end{pmatrix}$$

so that

$$\mathbf{AG} = \begin{pmatrix} 1 & 0 \\ 0 & \phi_2 \end{pmatrix} \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ \phi_2 & 0 \end{pmatrix}$$

which has the stated– symmetric– form.

General case: This is easily verified simply by working through the matrix product entry-by-entry. And $\mathbf{AG}$ is evidently symmetric.

(b) Show or deduce that:

i. For a proper AR(p) model in which $\phi_p \neq 0$, then $|\mathbf{A}| \neq 0$ so that $\mathbf{A}$ is non-singular.

Special case of $p = 2$: In this case

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & \phi_2 \end{pmatrix}.$$

Since $\phi_2 \neq 0$ then $|A| = \phi_2 \neq 0$ so $\mathbf{A}^{-1}$ exists.

General case: From the form of $\mathbf{A}$, is is clear that $|\mathbf{A}| = \phi_p^{p-1}$ which is non-zero hence $\mathbf{A}$ is non-singular.

ii. $\mathbf{AGA}^{-1} = \mathbf{G}'$ (*you can do this* without *trying to invert* $\mathbf{A}$).

Special case of $p = 2$: Here, directly,

$$\mathbf{A}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \phi_2^{-1} \end{pmatrix}.$$

Then direct matrix multiplication gives the stated result. *Or*, follow the logic– below– for the case of general p that uses symmetry argument and does not need you to invert any matrices.

General case: Since $\mathbf{AG}$ is symmetric, then $\mathbf{AG} = (\mathbf{AG})' = \mathbf{G}'\mathbf{A}' = \mathbf{G}'\mathbf{A}$ as $\mathbf{A}$ is also symmetric. Thus, since $\mathbf{A}$ is non-singular, we deduce $\mathbf{AGA}^{-1} = \mathbf{G}'$.

iii. $\mathbf{AF} = \mathbf{F}$ and, as a result, $\mathbf{F}' = \mathbf{F}'\mathbf{A}^{-1}$.

This is immediate from the forms of $\mathbf{F}$ and $\mathbf{A}$.

(c) Hence show that an equivalent state space AR(p) form is given by $y_t = \mathbf{F}'\mathbf{z}_t$ and $\mathbf{z}_t = \mathbf{G}'\mathbf{z}_{t-1} + \mathbf{F}\epsilon_t$ based on a new $p \times 1$ state vector $\mathbf{z}_t = \mathbf{A}\mathbf{x}_t$ and where the state evolution matrix is $\mathbf{G}'$, i.e.,

$$\mathbf{G}' = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \phi_{p-1} & 0 & 0 & \cdots & 1 \\ \phi_p & 0 & 0 & \cdots & 0 \end{pmatrix}$$

From the initial state space form, we see that

$$\mathbf{z}_t = \mathbf{A}\mathbf{x}_t = \mathbf{AG}\mathbf{x}_{t-1} + \mathbf{AF}\epsilon_t = \mathbf{AGA}^{-1}\mathbf{A}\mathbf{x}_{t-1} + \mathbf{AF}\epsilon_t = \mathbf{G}'\mathbf{z}_{t-1} + \mathbf{F}\epsilon_t$$

using the results above. Further, $y_t = \mathbf{F}'\mathbf{x}_t = \mathbf{F}'\mathbf{A}^{-1}\mathbf{z}_t = \mathbf{F}'\mathbf{z}_t$. The result follows.

(d) What is the interpretation of the elements of the transformed state vector $\mathbf{z}_t$?

Special case of $p = 2$: Here $\mathbf{z}_t' = (y_t, \phi_2 y_{t-1})$ so the first element is just as in the original model– the current value of the series– while the second element is the *contribution to the linear regression* for y_t from the lag-2 value.

General case: The lead element of the transformed state vector $\mathbf{z}_t$ is just y_t as in the original model. The later values are linear combinations of lagged y_{t-j} elements that contribute to the linear (auto-)regression for y_t. Not so interpretable, at all, as the original state space formulation, right? But the latter representation is nevertheless common in some areas of application.

One unstated point that we have reviewed in class: *Any* $p \times p$ matrix $\mathbf{A}$ such that $\mathbf{AGA}^{-1}$ has the same eigenvalues as $\mathbf{G}$ could be used to develop this analysis– transforming to a new state vector $\mathbf{z}_t = \mathbf{A}\mathbf{x}_t$ without changing the form of the model. In this example, $\mathbf{G}'$ is the result– and of course has the same eigenvalues as $\mathbf{G}$. Generally, it is wise to choose the "simplest" form and one that has easiest interpretation from such a class of "similar" models.

Bonus points for developing some of the theory in the general case.