1. Consider the stationary model $x_t \leftarrow AR(1|(\phi, v))$ with $s = 1$.

   - Simulate series of length $n = 100$ from the distribution of $x_{1:n}$ conditional on an pre-initial value $x_0$ for the two cases $x_0 = 0$ and $x_0 = 10$. Repeat this a few times with different values of $\phi$ (large, positive and negative in particular). Describe and interpret the resulting realizations, in comparison to realizations from the joint distribution $p(x_{1:n})$. (Matlab code should be part of your write-up).
   - In terms of $x_0, \phi, v$, give expressions for $E(x_t|x_0)$ and $V(x_t|x_0)$ for any $t > 0$. Interpret what happens to these values as $t \to \infty$.

2. In the stationary model $x_t \leftarrow AR(1|(\phi, v))$, we know $(x_t|x_{t-1}) \sim N(\phi x_{t-1}, v)$ and $(x_{t+1}|x_t) \sim N(\phi x_t, v)$; and of course the first-order Markov structure implies that $(x_{t+1}|x_t) \perp \perp x_{t-1}$, for all $t$.

   - What is the distribution of $(x_t|x_{t-1}, x_{t+1})$?
   - Define $x_{1:n}\backslash t = (x_1, \ldots, x_{t-1}, x_{t+1}, \ldots x_n)'$. What is the distribution of $(x_t|x_{1:n}\backslash t)$? (nb., there are at least two direct ways to answer this, both requiring no more than one or two lines of answer.)

3. Consider the model $x_t \leftarrow AR(1|(\phi, 1))$ with $|\phi| < 1$ but in which the innovations have a standard Cauchy distribution rather than a normal (and are still independent).

   * One easy way to simulate Cauchy random variables is to use the definition: $z$ has a standard Cauchy distribution if $z = v/u$ where $v, u$ are independent standard normal variates. A trivial change in the example Matlab code is to use $z = \text{randn}(n,1)/\text{randn}(n,1)$. There are other ways as we shall see.

   - Simulate sample paths from this model taking values of $\phi$ in $0.9, -0.9, 0.1, -0.1$ and for $n = 1000$. Repeat the simulations a few times to get the flavour of variation among typical sample paths. In addition to examining plots of the time series, look at lag 1 scatter plots of $x_t$ versus $x_{t-1}$.
   - Briefly describe the main ways in which these realizations differ from those generated by Gaussian models.
   - Is the process first-order Markov?
   - Is it a linear process?
   - Is it stationary?
   - Is it weakly stationary?
   - Do you think the process is time reversible?
4. Refer back to Exercise 0.3 of the STA 214 course Orientation Note and Exercises. In the special case of that example for which \( k = 1 \), the two random variables couple to define a Cauchy distribution: changing notation, if \( p(\epsilon, \lambda) = p(\epsilon | \lambda) p(\lambda) \) where \( (\epsilon | \lambda) \sim N(0, 1 / \lambda) \) and \( \lambda \sim Ga(1/2, 1/2) \), then that exercise leads to the implied marginal density \( p(\epsilon) = 1 / \{ \pi (1 + \epsilon^2) \} \), the density of the standard Cauchy distribution. With this in mind:

- A very modest change has \( (\epsilon | \lambda) \sim N(0, v / \lambda) \) for some fixed \( v > 0 \). How does this change \( p(\epsilon) \)?
- In the AR(1) model, suppose that, at each time \( t \), the innovation error \( \epsilon_t \) has a slightly modified distribution: in place of \( N(0, v) \), \( \epsilon_t \) is now distributed as \( N(0, v / \lambda_t) \) conditional upon some positive “weight” \( \lambda_t \). As a model this allows for a much larger innovation at a time \( t \) than \( N(0, v) \) generates if \( \lambda_t \) is small, and for smaller innovations when weights are large. Suppose that we model the weights as a random sample from a gamma distribution, specifically \( \lambda_t \sim Ga(1/2, 1/2) \) independently. What is the implied distribution of each \( \epsilon_t \)?
- Describe how this provides an alternative approach to simulating the AR(1) model with Cauchy innovations of Question 3 above.

5. In the stationary AR(1) plus noise model, we observe \( y_t \) where

\[
\begin{align*}
y_t &= x_t + \nu_t \\
x_t &\leftarrow AR(1|\theta)
\end{align*}
\]

where \( \nu_t \sim N(0, w) \) and with \( \nu_t \perp \nu_s \) and \( \nu_t \perp \epsilon_s \) for all \( t, s \). We know that \( y_t \sim N(0, q) \).

- Find an expression for the variance \( q \) in terms of the parameters \( \theta = (\phi, v) \) and \( w \).
- Show that we can write \( y_t = \phi y_{t-1} + \delta_t \) where \( \delta_t \) is a linear function of \( \epsilon_t, \nu_t \) and \( \nu_{t-1} \).
- Obviously each \( \delta_t \) is normally distributed. Find the mean and variance of \( \delta_t \). What is the correlation between \( \delta_t \) and \( \delta_{t-1} \)?
- Is the \( y_t \) process Markovian? Why, or why not?

6. Generate a reasonably large random sample from the \( \chi^2_1 \) distribution and look at histograms of the implied values of \( \nu_t = \log(\kappa_t) / 2 \) where \( \kappa_t \sim \chi^2_1 \) as in the non-linear stochastic volatility model. To further explore the shape (including asymmetry) in this distribution, look at qqplots (against the standard normal distribution) to get more insight into how the distribution of \( \nu_t \) differs from normal, especially in the tails. Briefly describe the differences.

7. Intellectual bonus question (not expected/required for credit): Can you find the inverse of the AR(1) variance matrix

\[
\Phi_n = \begin{pmatrix}
1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\
\phi & 1 & \phi & \cdots & \phi^{n-2} \\
\phi^2 & \phi & 1 & \cdots & \phi^{n-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & 1
\end{pmatrix}
\]

by direct linear algebra methods? (There are other, easier ways).
Solutions to Homework #1

1. Sample paths starting at a value of $x_0$ that is not sampled, or distributed, as $N(0, s)$ do not represent samples from the full joint distribution. However, the short-memory of an AR(1) process means that the effects of such chosen initial values decay exponentially ($|\phi|^t$) so that, after some time steps, the realization of the process converges; eventually for $t$ large enough, $x_{t:(t+n)}$ does indeed represent a sample from $p(x_t, x_{t+1}, \ldots, x_{t+n})$. The effect of initial values chosen like this is longer for larger $|\phi|$. Generating sample paths from $x_0 = 0$ leads to trajectories that are really almost immediately indistinguishable from those generated from a random initial value $x_0 \sim N(0, s)$ since this distribution is centred at 0. Starting at $x_0 = 10$ when $s = 1$, however, starts the process at what is a very unlikely value; the process then quickly moves towards 0, either decreasing (for cases with $\phi > 0$) or oscillating (when $\phi < 0$) and - once values move into the plausible range of the $N(0, s)$ stationary marginal distribution.

By iterating expectations, $E(x_t|x_0) = \phi^t x_0$ and

\[ V(x_t|x_0) = v(1 + \phi^2 + \cdots + \phi^{t-1}) = v(1 - \phi^{2(t-1)})/(1 - \phi^2) = s(1 - \phi^{2(t-1)}). \]

So as $t$ increases, $E(x_t|x_0) \to 0$ and $V(x_t|x_0) \to s$ at an exponential rate, and the distribution $p(x_t|x_0)$ – that is of course normal with these moments – converges quickly to the stationary marginal distribution for each $x_t$.

2. By Bayes’ theorem $p(x_t|x_{t-1}, x_{t+1}) \propto p(x_t|x_{t-1})p(x_{t+1}|x_t, x_{t-1})$; from the model, the two terms here are $(x_t|x_{t-1}) \sim N(\phi x_{t-1}, v)$ and $(x_{t+1}|x_t, x_{t-1}) \sim N(\phi x_t, v)$. So

\[ -2v \log(p(x_t|x_{t-1}, x_{t+1})) = c + (x_{t+1} - \phi x_t)^2 + (x_t - \phi x_{t-1})^2 \]

which, on completing the quadratic, is a constant plus $(1 + \phi^2)(x_t - a_t)^2$ with

\[ a_t = (x_{t+1} + x_{t-1})/(1 + \phi^2). \]

So the distribution of $x_t$ conditional on its two neighbours is

\[ (x_t|x_{t-1}, x_{t+1}) \sim N(a_t, v/(1 + \phi^2)). \]

3. Cauchy innovations are, infrequently, very large in absolute value due to the heavy tails of the distribution. Occasional values are very, very much more extreme than those from normal models, and the effect on the AR(1) process is very marked when such values arise - the process shoots way out to values far from zero. In cases of high $|\phi|$, the process takes more time to return to regions nearer 0 than otherwise.
The AR(1) model is of course first-order Markov, as that property is not affected by the assumed form of the innovation distribution.

The process is linear. Again, distributional form for the innovations does not impact the model structure. But note: In the theoretical stochastic process literatures, linear process is often restricted to linear processes such as we have here but in which the innovations have (at least) finite variance. The Cauchy model falls outside that narrower definition; much of the theory of linear processes is restricted to that narrower set.

The process is indeed stationary. To prove this we need to show that the marginal distribution of any \( x_t \) exists and is the same for all \( t \), which can be done a number of ways (and will be later on), and then the result follows by inspection.

Since the Cauchy distribution has no mean or variance, weak stationarity has no meaning for these processes. This shows a weakness of the definition of weak stationarity relative to the global (strong) stationarity concept.

Sample paths simulated will suggest the process is not time reversible. This is best seen in cases of larger positive values of \( \phi \), when a large, positive innovation shocks the series to high values, and there is then a decay back towards zero that exhibits a “hysteresis” effect - the shock upwards is followed by a slower decay, which is a phenomenon that is not symmetric with respect to reversal of the time axis. The theory of reversible linear processes is heavily dependent on finite variance innovations, and this is an example outside that class.

\[ p(\epsilon) \propto 1/\{1 + \epsilon^2/v\} \] so that \( \epsilon/\sqrt{v} \) is standard Cauchy, or \( \epsilon \sim C(0, v) \).

\( \epsilon_t \sim C(0, v) \), so we have the Cauchy innovation-driven AR(1) model.

Generate a sample \( \epsilon_t \) by: (a) first sample a scale factor \( \lambda_t \sim Ga(1/2, 1/2) \), then (b) based on this value of \( \lambda_t \), sample \( \epsilon_t \) from the normal \( N(0, v/\lambda_t) \).

\[ q = V(y_t) = V(x_t + \nu_t) = s + w \] using \( x_t \perp \nu_t \) so that their covariance is 0.

Substitute \( x_{t-1} = y_{t-1} - \nu_{t-1} \) in the AR(1) model expression to get \( y_t = \phi(y_{t-1} - \nu_{t-1}) + \epsilon_t + \nu_t \) which reduces to the form \( (1 - \phi B)y_t = (1 - \phi B)\nu_t + \epsilon_t \), or \( (1 - \phi B)y_t = \delta_t \) where \( \delta_t = (1 - \phi B)\nu_t + \epsilon_t \).

\[ E(\delta_t) = 0 \] immediately. Also, using independence and the implied lack of covariance between the error terms, \( V(\delta_t) = V(\nu_t) + \phi^2V(\nu_{t-1}) + V(\epsilon_t) = (1 + \phi^2)w + v \).

Further, since the covariance between any pair of the error terms is zero, the covariance at lag 1 is just that of the only term in common,

\[ C(\delta_t, \delta_{t-1}) = E(-\phi\nu_{t-1}^2) = -\phi w, \]

so that the lag 1 correlation is \(-\phi/(1 + \phi^2 + v/w)\).
Notice that the observational error terms have the effect of modifying the original AR(1) process so that the new innovations, $\delta_t$, are correlated negatively. This is an example of an ARMA process - the model here implied for $y_t$ is a stationary, ARMA(1,1) model, with $\delta_t$ now contributing the MA (moving-average) term of order 1.

- $y_t$ is not a Markovian process (unless $w = 0$.) The reason is that the correlation in the $\delta_t$ sequence means that all past $y_s$ values provide information about $y_t$ now indirectly through the $\delta$ series.

6. The distribution of $\nu_t = \log(\kappa_t)/2$ where $\kappa_t \sim \chi^2_1$ has some asymmetry. Relative to the normal, the distribution is heavier tailed on negative values and lighter tailed on positive values. As a model for errors or innovations in a process, this will therefore generate relative more and larger negative inputs than a normal model.

Total score out of: 26pt