1. **AR(2) Processes**

A stationary stochastic process $\{X_t : t = 0, \pm 1, \pm 2, ...\}$ satisfies the relationship

$$X_t = \mu + 0.8(X_{t-1} - \mu) - 0.4(X_{t-2} - \mu) + W_t,$$

where $\{W_t : t = 0, \pm 1, \pm 2, ...\}$ is a sequence of independent, zero-mean normal random variables with common variance σ^2.

(a) Is the model causal? Justify your answer.

(b) Is the model invertible? Justify your answer.

(c) Calculate $\rho(1)$ and $\rho(2)$, the autocorrelation functions of $\{X_t\}$ evaluated at lag 1 and 2 respectively.

(d) What are ϕ_{11} and ϕ_{22}, the partial autocorrelation function at lag 1 and 2 respectively? (Hint: This problem requires no computation.)

(e) What are ϕ_{kk}, the partial autocorrelation function at lag k, for $k \geq 3$?

In a realistic setting, we *observe* the series $(x_1, ..., x_n)$ for a finite n. Suppose $n = 300$ and the process is indeed a stationary AR(2) process. That is,

$$x_t = \mu + \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + w_t,$$

where w_t are i.i.d. *observations* of Normal$(0, \sigma^2)$ random variables $\{W_t\}$, where ϕ_1, ϕ_2, μ, and σ^2 are unknown parameters to be estimated.

(f) Show how to reparametrize the model in the intercept form

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t.$$

(g) Write down the conditional likelihood function (conditioning on x_1 and x_2) of $(\alpha, \phi_1, \phi_2, \sigma^2)$.

(h) The conditional MLE of (α, ϕ_1, ϕ_2) can be expressed as

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z},$$

for some matrix $\mathbf{W}$ and some vector $\mathbf{Z}$ whose entries are based on $(x_1, ..., x_n)$. Write out what $\mathbf{W}$ and $\mathbf{Z}$ are.
Hint: Think about linear regression!

Suppose the conditional MLEs are found to be $(\hat{\alpha}, \hat{\phi}_1, \hat{\phi}_2, \hat{\sigma}^2) = (0, 3/4, -1/8, 9)$.

(i) Based on the MLEs, find the (estimated) causal representation of x_t.

(j) Given $x_{98} = 8, x_{99} = 16, x_{100} = 12$, taking the MLEs as the true parameter values, construct a 95% forecast interval for x_{101}.

2. **Asymptotic distributions of parameter estimates**

Let n be the number of observations and consider conditional least-squares estimates of the AR coefficients in a causal AR process. For AR(1) $x_t = \phi x_{t-1} + w_t$,

$$\hat{\phi} \overset{\text{approx.}}{\sim} N\left(\phi, \frac{1-\phi^2}{n}\right)$$

and for AR(2) $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$,

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} \overset{\text{approx.}}{\sim} N\left(\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}, \frac{1}{n}\begin{bmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{bmatrix}\right).$$

In this exercise we will verify some of the insights from these formulae via simulations. Execute in **R** the steps described in the box below, and answer the questions that follow.

> For $i = 1, ..., 1000$:
>
> (i) Simulate an AR(1) process with $x_t = 0.8x_{t-1} + w_t$, (w_t) i.i.d. $N(0,1)$ with $n = 300$.
>
> (ii) Use `arima()` to fit an AR(1) model. Let $\hat{\phi}^{(i)}_{1,AR(1)}$ be the estimate of ϕ_1.
>
> (iii) Use `arima()` to fit an AR(2) model. Let $\hat{\phi}^{(i)}_{1,AR(2)}$ be the estimate of ϕ_1.

(a) Compute the mean and the variance of $(\hat{\phi}^{(1)}_{1,AR(1)}, ...\hat{\phi}^{(1000)}_{1,AR(1)})$. Plot the histogram of $(\hat{\phi}^{(1)}_{1,AR(1)}, ...\hat{\phi}^{(1000)}_{1,AR(1)})$. Are these consistent with the asymptotic results?

(b) Compute the mean and the variance of $(\hat{\phi}^{(1)}_{1,AR(2)}, ...\hat{\phi}^{(1000)}_{1,AR(2)})$. Plot the histogram of $(\hat{\phi}^{(1)}_{1,AR(1)}, ...\hat{\phi}^{(1000)}_{1,AR(1)})$. Are these consistent with the asymptotic results?

(c) Plot the density curves of $(\hat{\phi}^{(1)}_{1,AR(1)}, ...\hat{\phi}^{(1000)}_{1,AR(1)})$ and $(\hat{\phi}^{(1)}_{1,AR(2)}, ...\hat{\phi}^{(1000)}_{1,AR(2)})$ in a single plot. What do you notice? What is the issue of fitting an AR(2) model to an AR(1) process?