1 Basic Concepts

- **Mean Function:** $\mu_t = \mathbb{E}[X_t]$
- **Autocovariance Function:** $\gamma(h) = \text{Cov}(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)]$
- **Autocorrelation Function (ACF):** $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$
- **Partial Autocorrelation Function (PACF):** Direct correlation between X_t and X_{t-h} after removing intermediate effects. PACF connects to AR and ACF connects to MA.

2 Stationarity

- **Strict Stationarity:** Joint distribution of $(X_{t_1}, \ldots, X_{t_k})$ is invariant to time shifts.
- **Weak Stationarity:**
  - $\mathbb{E}[X_t] = \mu$ (constant)
  - $\text{Var}(X_t) = \sigma^2$ (constant)
  - $\text{Cov}(X_t, X_{t+h})$ depends only on h

3 AR, MA, and ARMA Models

Autoregressive Model (AR(p))

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$$

- Stationary if roots of characteristic equation lie outside the unit circle

4 Yule-Walker Equations (for AR(p))

$$\gamma(k) = \sum_{j=1}^{p} \phi_j \gamma(k-j), \quad \text{for } k = 1, 2, \ldots, p$$

- Used to estimate parameters of an AR model
- $\gamma(k)$ is the autocovariance at lag k
- ϕ_j are the AR coefficients

Moving Average Model (MA(q))

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

- Invertible if roots of MA polynomial lie outside the unit circle

ARMA(p, q)

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

5 ARIMA(p, d, q)

- Differencing: $\nabla X_t = X_t - X_{t-1}, \quad \nabla^2 X_t = \nabla(\nabla X_t)$
- X_t is ARIMA(p, d, q) if $\nabla^d X_t \sim ARMA(p, q)$
- $r_{1,1} = r_1 = \rho(1)$
- $r_{k,k} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j}$
- $r_{k,j} = r_{k-1,j} - r_{k,k} r_{k-1,k-j}$

6 Forecasting (for AR(1))

$$X_t = \mu + \phi(X_{t-1} - \mu) + Z_t$$

$$\hat{X}_{t+h} = \mu + \phi^h (X_t - \mu)$$

7 Variance and Covariance Shortcuts

- $\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab\text{Cov}(X, Y)$
- $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

8 Model Properties

AR(1): $X_t = \phi X_{t-1} + Z_t$

- $\mathbb{E}[X_t] = 0$, $\text{Var}(X_t) = \frac{\sigma^2}{1-\phi^2}$ for $|\phi| < 1$
- ACF: $\rho(h) = \phi^h$
- PACF: cuts off after lag 1

AR(2): $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$

- $\text{Var}(X_t) = \frac{\sigma^2}{1-\phi_1^2-\phi_2^2-2\phi_1\phi_2\rho(1)}$
- ACF:
$$\rho(1) = \frac{\theta_1 + \theta_1 \rho(1)}{1 - \theta_2}, \quad \rho(2) = \frac{\theta_1 \rho(1) + \theta_2}{1 - \theta_2}$$
- PACF: cuts off after lag 2

MA(1): $X_t = Z_t + \theta Z_{t-1}$

- ACF: $\rho(1) = \frac{\theta}{1+\theta^2}$, cuts off after lag 1
- PACF: tails off

MA(2): $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$

- ACF:
$$\rho(1) = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho(2) = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho(k) = 0 \ \forall \ k \geq 3$$
- PACF: tails off

9 Forecasting

AR(1) Forecast

$$\hat{X}_{t+h} = \mu + \phi^h (X_t - \mu)$$

$$\text{Var}(\hat{X}_{t+h}) = \sigma^2 \cdot \frac{1 - \phi^{2h}}{1 - \phi^2}$$

10 Model Identification (ACF & PACF)

Model	ACF	PACF
AR(p)	Tails off	Cuts off at p
MA(q)	Cuts off at q	Tails off
ARMA(p,q)	Tails off	Tails off

11 Unit Root Test (Dickey-Fuller)

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \epsilon_t$$

Test $H_0 : \gamma = 0$ (unit root). Reject $\Rightarrow$ stationary.

12 Lag Operator Rules

- $BX_t = X_{t-1}, \quad B^k X_t = X_{t-k}$
- $\nabla X_t = (1 - B)X_t$
- $\nabla^d X_t = (1 - B)^d X_t$
- $\nabla_L X_t = (1 - B^L)X_t$

13 SARIMA(p, d, q)x(P, D, Q)

The general SARIMA model is given by:

$$\Phi_P(B^s)\phi_p(B)\nabla^d \nabla_s^D X_t = \Theta_Q(B^s)\theta_q(B)Z_t$$

- B: backshift operator, $BX_t = X_{t-1}$
- $\nabla = 1 - B$: non-seasonal difference operator
- $\nabla_s = 1 - B^s$: seasonal difference operator
- $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$: non-seasonal AR
- $\theta_q(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$: non-seasonal MA
- $\Phi_P(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps}$: seasonal AR
- $\Theta_Q(B^s) = 1 + \Theta_1 B^s + \cdots + \Theta_Q B^{Qs}$: seasonal MA
- $Z_t \sim WN(0, \sigma^2)$: white noise

Given:

$X_t = 0.3X_{t-1} + X_{t-4} - 0.3X_{t-5} + Z_t - 0.3Z_{t-1} - 0.5Z_{t-4} + 0.15Z_{t-5}$

We rewrite this in terms of backshift operator B:

$$(1 - 0.3B - B^4 + 0.3B^5)X_t = (1 - 0.3B - 0.5B^4 + 0.15B^5)Z_t$$

This can be factored as:

$$\phi(B)\Phi(B^s) = (1 - 0.3B)(1 - B^4), \quad \theta(B)\Theta(B^s) = (1 - 0.3B)(1 - 0.5B^4)$$

This corresponds to a seasonal ARIMA model with:

- B: backshift operator, $BX_t = X_{t-1}$
- $\nabla = 1 - B$: non-seasonal difference operator
- $\nabla_s = 1 - B^4$: seasonal difference operator (since lag 4 appears)
- $\phi_p(B) = 1 - 0.3B$: non-seasonal AR(1)
- $\theta_q(B) = 1 - 0.3B$: non-seasonal MA(1)
- $\Phi_P(B^s) = 1 - B^4$: seasonal AR(1), with $s = 4$
- $\Theta_Q(B^s) = 1 - 0.5B^4$: seasonal MA(1), with $s = 4$
- $Z_t \sim WN(0, \sigma^2)$: white noise

So the full model is:

$$\text{SARIMA}(p = 1, d = 0, q = 1) \times (P = 1, D = 1, Q = 1)_{s=4}$$

14 Important extras

- If $e_t \sim \mathcal{N}(0, \sigma^2)$, then:
$$\mathbb{E}[e_t] = 0, \quad \mathbb{E}[e_t^2] = Var(e_t) = \sigma^2, \quad \mathbb{E}[e_t^3] = 0, \quad \mathbb{E}[e_t^4] = 3\sigma^4$$
- Covariance with powers:
$$\text{Cov}(e_t, e_t^2) = \mathbb{E}[e_t^3] = 0$$
$$\text{Cov}(ae_t, be_t^2) = ab \cdot \left(\mathbb{E}[e_t^3] - \mathbb{E}[e_t]\mathbb{E}[e_t^2]\right) = 0$$

15 Other useful formulas

$$\cos(x)\cos(y) = \frac{1}{2}[\cos(x-y) + \cos(x+y)]$$

$$\sin(x)\sin(y) = \frac{1}{2}[\cos(x-y) - \cos(x+y)]$$

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \quad \text{for } |x| < 1$$

16 ACF and PACF estimations

Let $\{Z_t\}$ be white noise with mean zero and variance one.

(a) Find the ACFs and PACFs for the model in (i), and the ACFs for the model in (ii).

(i)
$$X_t - 0.3X_{t-2} = Z_t$$

(ii)
$$X_t = 0.5X_{t-1} + Z_t + 0.7Z_{t-1} + 0.6Z_{t-2}$$

(i) Multiplying X_{t-k} and taking expectation we obtain:

$$E(X_t X_{t-k}) = 0.3E(X_{t-2}X_{t-k}) + E(Z_t X_{t-k}),$$

which is equivalent to:

$$\gamma(k) = 0.3\gamma(k-2), \quad k > 0.$$

This implies that:

$$\rho(2k) = 0.3^k, \quad \rho(2k-1) = 0, \quad j = 1, 2, \ldots$$

Its PACF is given by:

$$r_{11} = 0, \quad r_{22} = 0.3, \quad r_{kk} = 0, \quad k > 2.$$

(ii) Taking expectation on both sides of the model:

$$X_t = 0.5X_{t-1} + Z_t + 0.7Z_{t-1} + 0.6Z_{t-2}$$

yields $\mathbb{E}[X_t] = 0$.

Multiplying both sides by X_{t-k} and taking expectation:

$$\begin{aligned} E(X_t X_{t-k}) &= 0.5E(X_{t-1}X_{t-k}) + E(Z_t X_{t-k}) \\ &+ 0.7E(Z_{t-1}X_{t-k}) + 0.6E(Z_{t-2}X_{t-k}) \end{aligned} \tag{1}$$

When $k > 2$, we have:
$$\rho_k = 0.5\rho_{k-1} \tag{2}$$

When $k = 2$, we have:

$$\rho_2 = 0.5\rho_1 + 0.6\frac{E(Z_{t-2}X_{t-2})}{\gamma_0} \tag{3}$$

When $k = 1$, we have:

$$\rho_1 = 0.5\rho_0 + 0.7\frac{E(Z_{t-1}X_{t-1})}{\gamma_0} + 0.6\frac{E(Z_{t-2}X_{t-1})}{\gamma_0} \tag{4}$$

When $k = 0$, equation (1) becomes:

$$\begin{aligned} \gamma_0 &= 0.5\gamma_1 + E(Z_t X_t) + 0.7E(Z_{t-1}X_t) + 0.6E(Z_{t-2}X_t) \\ &= 0.5\gamma_1 + 1 + 0.7 \cdot 1.2 + 0.6 \cdot 1.2 \\ &= 0.5\gamma_1 + 2.56 \end{aligned}$$

$$1 = 0.5\rho_1 + \frac{2.56}{\gamma_0} \tag{5}$$

From (4) and (5) we obtain:

$$\rho_1 = \frac{3.6}{4.36} \approx 0.82568$$

Now compute required expectations:

$$E(Z_{t-1}X_{t-1}) = E(X_{t-1}Z_{t-1}) =$$

$$0.5E(X_{t-2}Z_{t-1}) + E(Z_{t-1}^2) + 0.7E(Z_{t-2}Z_{t-1}) + 0.6E(Z_{t-3}Z_{t-1}) = 1$$

$$E(Z_{t-2}X_{t-1}) =$$

$$0.5E(Z_{t-2}X_{t-2}) + E(Z_{t-2}Z_{t-1}) + 0.7E(Z_{t-2}^2) + 0.6E(Z_{t-2}Z_{t-3}) = 0.7$$

$$E(Z_{t-2}X_t) = 0.5E(Z_{t-2}X_{t-1}) + 0.6E(Z_{t-2}^2) = 0.5 \cdot 1.2 + 0.6 = 1.2$$

From (4) and (3), we now find:

$$\gamma(0) = 4.36, \quad \rho_2 = 0.5 \cdot 0.82568 + \frac{0.6 \cdot 1.2}{4.36} = 0.55$$

$$\rho_k = 0.5\rho_{k-1}, \quad k > 2$$

17 Sample PACF Analysis and Seasonal ARIMA Classification

Given: From 100 time series observations, we calculate its sample PACFs:
$r_{11} = 0.6, \quad r_{22} = 0.5, \quad r_{33} = 0.1, \quad$ sample mean $\bar{x} = 1, \quad$ sample variance $= 5$

(a) (i) Suggest an appropriate model and justify it.

(ii) Estimate the parameters involved in the model and write down the explicit expression.

Solution:

(i) The critical value is $\frac{1.96}{\sqrt{100}} = 0.196$. Since $r_{11}, r_{22} > 0.196$ and $r_{33} < 0.196$, an AR(2) model is appropriate, as PACF cuts off after lag 2.

(ii) Consider the AR(2) model:
$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$$

From the Yule-Walker equations:
$$\rho_1 = \phi_1 + \phi_2\rho_1, \quad \text{where } \rho_1 = r_{11} = 0.6, \quad \phi_2 = r_{22} = 0.5$$

Solving gives:
$$\phi_1 = \rho_1 - \phi_2\rho_1 = 0.6 - 0.5 \cdot 0.6 = 0.3$$

The drift term estimate:
$$\hat{\delta} = (1 - \phi_1 - \phi_2)\bar{x} = (1 - 0.3 - 0.5)(1) = 0.2$$

Variance computation:
$$\begin{aligned} \sigma^2 = Var(Z_t) &= (1 - \phi_1^2 - \phi_2^2)Var(X_t) - 2\phi_1\phi_2\rho_1 \\ &= (1 - 0.3^2 - 0.5^2)(5) - 2 \cdot 0.3 \cdot 0.5 \cdot 0.6 = 3.12 \end{aligned}$$

(b) Suppose:
$$X_t = 0.3X_{t-1} + X_{t-4} - 0.3X_{t-5} + Z_t - 0.3Z_{t-1} - 0.5Z_{t-4} + 0.15Z_{t-5}$$

Classify the model as a seasonal ARIMA model and write down the explicit values of p, d, q, P, D, Q and the seasonal period s.

Solution:

Characteristic polynomials:

$$\phi(z) = 1 - 0.3z - z^4 + 0.3z^5 = (1 - 0.3z)(1 - z^4)$$

$$\theta(z) = 1 - 0.3z - 0.5z^4 + 0.15z^5 = (1 - 0.3z)(1 - 0.5z^4)$$

Hence, the seasonal ARIMA classification is:
$$(p, d, q)(P, D, Q)_s = (1, 0, 1)(0, 1, 1)_4$$

SARIMA Model Identification Cheat Sheet

Given: Monthly accidental deaths in the US (1973–1978). Assume plots are provided.

(i) **Component Identification (Fig 1):**
- Trend component (upward/downward over time)
- Seasonal component (repeats every 12 months)
- Irregular/random noise

(ii) **Interpretation of Differenced Plots (Figs 2 & 3):**
- Fig 2: First differencing reduces trend but seasonal pattern remains.
- Fig 3: Seasonal differencing reduces seasonality but trend remains.
- Both regular and seasonal differencing required: $d = 1, D = 1$

(iii) **R Statements:**
- `acf(diff(diff(death, 12), 1), 36)`: ACF of data differenced at lag 12 (seasonal) then 1 (regular), up to lag 36.
- `pacf(diff(diff(death, 12), 1), 36)`: PACF of the same series.

(iv) **Model Suggestion (based on ACF/PACF – Fig 4):**
- ACF: spikes at lag 1 and seasonal lag MA(1) and seasonal MA(1)
- PACF: spike at lag 1 AR(1)
- Common SARIMA candidates:

$$\text{(I) SARIMA}(0,1,1) \times (0,1,1)_{12}$$
$$\text{(II) SARIMA}(1,1,0) \times (1,1,0)_{12}$$
$$\text{(III) SARIMA}(0,1,1) \times (1,1,0)_{12}$$
$$\text{(IV) SARIMA}(1,1,0) \times (0,1,1)_{12}$$

(v) **Model Diagnostics (Fig 5):**
- Residuals: appear randomly scattered (no trend/seasonality)
- ACF of residuals: no significant spikes
- Ljung-Box p-values > 0.05 ==> residuals are uncorrelated
- ==> Model is adequate

Figure 1: The plot of the original data.

Figure 2: the plot for taking the first differenced data

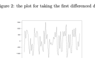

Figure 3: the plot for taking the seasonal differenced data

Figure 4: ACF and PACF plots for taking both nonseasonal and seasonal differenced data

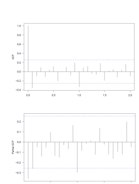

Figure 5: Diagnostic plots