Covariance, Correlation Formulas

$$\mathrm{Cov}(a_0 + \sum_{j=1}^{p} a_j X_j, b_0 + \sum_{k=1}^{q} b_k Y_k) = \sum_{j=1}^{p}\sum_{k=1}^{1} a_j b_k \mathrm{Cov}(X_j, Y_k)$$

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}, \quad X, Y \in L_2$$

Times Series Data Features

- Trend: mean/tendency.
- Periodicity: repetition in pattern.
- Seasonality: periodicity w/ known period.
- Heteroskedasticity: non-constant variance.
- Dependence: successive observations are similar/dissimilar.
- Other: missing data, structural breaks, outliers.

Stationarity

- Strong (SS): joint pdf/pmf invariant w/ time.
- Weak (WS): $1^{\mathrm{st}}, 2^{\mathrm{nd}}$ moments invariant w/ time.
- SS + Var $< \infty \implies$ WS.
- WS + jointly Guassian $\implies$ SS.

White Noise (WN)

$\{ Z_t \} \sim WN(0, \sigma^2)$ means that $\mathrm{E}(Z_t) = 0$, $\mathrm{Var}(Z_t) = \sigma^2\ \forall t \in \mathbb{Z}$ and $\mathrm{Cov}(Z_i, Z_j) = 0\ \forall i \neq j$.

Autocovariance Function (ACVF)

For WS $\{ X_t \}$, the ACVF $\gamma : \mathbb{Z} \to \mathbb{R}$ is $\gamma(h) = \mathrm{Cov}(X_t, X_{t+h})$.
Properties:

1. $\gamma(0) = \mathrm{Var}(X_t),\ \forall t \in \mathbb{Z}$,
2. By the Cauchy-Schwarz inequality,
$$|\gamma(h)| = \left|\mathrm{Cov}(X_t, X_{t+h})\right| \le \sqrt{\mathrm{Var}(X_t)\,\mathrm{Var}(X_{t+h})} = \gamma(0).$$
3. Even: $\gamma(-h) = \gamma(h),\ \forall h \in \mathbb{Z}$.
4. Non-negative definite: $\forall n,\ \mathbf{a} \in \mathbb{R}^n,\ \mathbf{t} \in \mathbb{Z}^n$,
$$\sum_{j=1}^{n}\sum_{i=1}^{n} a_i a_j \gamma(t_i - t_j) \ge 0.$$

Estimate by sample ACVF:

$$\hat\gamma(h) = \frac{1}{n}\sum_{t=1}^{n-|h|} (x_t - \bar x_n)(x_{t+|h|} - \bar x_n),\ \forall |h| < n.$$

Autocorrelation Function (ACF)

For WS $\{ X_t \}$, the ACF $\gamma : \mathbb{Z} \to [-1, 1]$ is $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$.
Estimate by sample ACF:

$$\hat\rho(h) = \frac{\hat\gamma(h)}{\hat\gamma(0)},\ \forall |h| < n.$$

Under mild conditions,

$$\hat{\boldsymbol\rho}_h = \left(\hat\rho(1), \hat\rho(2), \dots, \hat\rho(h)\right)' \approx \mathcal{N}\left(\boldsymbol\rho_h, \frac{1}{n}\mathbf{W}\right),$$

where the elements of $\mathbf{W} = [w_{ij}]_{i,j=1,\dots,h}$ are

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(k)\rho(i)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(k)\rho(j)\}.$$

Bartlett Bounds for sample ACF

When $\{ X_t \} \overset{\text{iid}}{\sim} (\mu, \sigma^2)$, $\rho(h) = 0\ \forall h \neq 0$. In the formula above,

$$w_{ij} = \sum_{k=1}^{\infty} \rho(k-i)\rho(k-j) = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{otherwise} \end{cases}$$

Then

$$\hat{\boldsymbol\rho}_h \approx \mathcal{N}\left(\mathbf{0}_h, \frac{1}{n}\mathbf{I}_{h\times h}\right)$$

and an approximate CI for $\rho(k)$ ("Bartlett Bounds") is

$$\hat\rho(k) \pm z_{1-\alpha/2}/\sqrt{n}.$$

Classical Decomposition

$$X_t = m_t + s_t + I_t,$$

where m_t allows for trend, s_t allows for seasonality, and I_t is the irregular/random part.

Removing Trend

- Linear filter (estimate m_t):
Pick $q \in \mathbb{Z}^+$ and a filter $\left\{ a_{-q}, \dots, a_0, \dots, a_1 \right\}$. Then $\hat m_t = \sum_{k=-q}^{q} a_k X_{t-k}$.
E.g., for a sample moving average, take $a_k = 1/(2q+1)$, so that (in absence of seasonality)
$$\hat m_t = \sum_{k=-q}^{q} \frac{X_{t-k}}{2q+1} = \underbrace{\sum_{k=-q}^{q} \frac{m_{t-k}}{2q+1}}_{q\downarrow \implies \downarrow \text{bias}} + \underbrace{\sum_{k=-q}^{q} \frac{I_{t-k}}{2q+1}}_{q\uparrow \implies \downarrow \text{variance}}.$$
- Exponential smoothing (estimate m_t):
Pick $a \in (0,1)$ and set
$$\begin{aligned}\hat m_1 &= X_1,\\ \hat m_t &= aX_t + (1-a)\hat m_{t-1}\\ &= aX_t + a(1-a)X_{t-1} + a(1-a)^2 X_{t-2} + \dots.\end{aligned}$$
- Differencing (eliminate m_t):
Apply difference operator $1 - B$, where B is the backshift operator: $B^k f(t) = f(t-k)$.
E.g., to kill linear trend $m_t = \alpha + \beta t$ use $(1-B)m_t = \beta$.
E.g., to kill quadratic trend $m_t = \alpha + \beta t + \gamma t^2$ use $(1-B)(1-B)m_t$.

Removing Seasonality (period d)

- Smoothing/filtering: $Y_t = \sum_{j=1}^{d-1} \frac{X_{t-j}}{d}$.
- Seasonal differencing: $1 - B^d$.
- Regression w/ dummy variables or trig polynomials: use to estimate s_t where $s_t = s_{t+d} = s_{t+2d} = \dots$ and $\sum_{j=1}^{d-1} s_{t-j} = 0$. Model as a linear combo of oscillating functions:
$$s_t = \sum_{j=1}^{\lfloor d/2 \rfloor} \left\{ a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t) \right\}$$
for frequencies $\lambda_j = \frac{2\pi j}{d}$.

Tests for WN

- Ljung-Box: reject $H_0 : \{ X_t \} \sim WN(0, \sigma^2)$ if $Q_{LB} > \chi^2_{h,1-\alpha}$, where
$$Q_{LB} = n(n+2)\sum_{k=1}^{h} \frac{\left[\hat\rho_X(k)\right]^2}{n-k} \overset{H_0}{\approx} \chi^2_n.$$
- McLeod-Li: reject $H_0 : \{ X_t \} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ if $Q_{ML} > \chi^2_{h,1-\alpha}$, where
$$Q_{ML} = n(n+2)\sum_{k=1}^{h} \frac{\left[\hat\rho_{X^2}(k)\right]^2}{n-k} \overset{H_0}{\approx} \chi^2_n.$$
- Tests for iid (randomness) based on ranks: let R_i = rank of X_i, $i = 1, \dots, n$.
  - Turning point: T = # of i where ranks jump up then down or vice versa.
  - Positive differences: S = # of i where $R_{i-1} < R_i$.
  - Positive pairs: P = # of (i,j) where $R_j > R_i$.

  For continuous iid data,
  - $\mathrm{E}(T) = \frac{2}{3}(n-2)$, $\quad \mathrm{Var}(T) = \frac{16n-29}{90}$.
  - $\mathrm{E}(S) = \frac{1}{2}(n-1)$, $\quad \mathrm{Var}(S) = \frac{n+1}{12}$.
  - $\mathrm{E}(T) = \frac{n(n-1)}{4}$, $\quad \mathrm{Var}(T) = \frac{n(n-1)(2n+5)}{72}$.

  Reject $H_0 : \{ X_t \}\ iid$ if $\left|\frac{\text{test stat - mean}}{\sqrt{\mathrm{Var}}}\right| > z_{1-\alpha/2}$.

Best MSE Prediction

To predict $X_{n+h} \mid X_1, \dots, X_n$ by minimizing MSE criterion

$$\mathrm{MSE}(\tilde X_{n+h}) = \mathrm{E}(X_{n+h} - \tilde X_{n+h})^2,$$

use

$$\hat X_{n+h} = \mathrm{E}(X_{n+h} \mid X_1, \dots, X_n).$$

Assume

$$\begin{pmatrix} X_{n+h} \\ X_n \\ \vdots \\ X_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}, \begin{pmatrix} \gamma(0) & \boldsymbol\gamma'(h) \\ \boldsymbol\gamma(h) & \boldsymbol\Gamma_n \end{pmatrix}\right),$$

where

$$\boldsymbol\gamma(h) = \mathrm{Cov}\left(X_{n+h}, \begin{pmatrix} X_n \\ \vdots \\ X_1 \end{pmatrix}\right) = \begin{pmatrix} \gamma(h) \\ \vdots \\ \gamma(n+h-1) \end{pmatrix}$$

$$\boldsymbol\Gamma_n = \mathrm{Var}\begin{pmatrix} X_n \\ \vdots \\ X_1 \end{pmatrix} = [\gamma(i-j)]_{i,j=1,\dots,n}.$$

Then the best MSE predictor (and best linear predictor) is

$$\hat X_{n+h} = \mu + \boldsymbol\gamma'(h)\boldsymbol\Gamma_n^{-1}\begin{pmatrix} X_n - \mu \\ \vdots \\ X_1 - \mu \end{pmatrix},$$

$$\mathrm{MSE}(\hat X_{n+h}) = \mathrm{E}(X_{n+h} - \hat X_{n+h})^2 = \gamma(0) - \boldsymbol\gamma'(h)\boldsymbol\Gamma_n^{-1}\boldsymbol\gamma(h).$$

Projection Theorem

Let $X_{n+h} \in L_2 = \{$ all r.v.'s with Var $< \infty \}$ and $\mathcal{M} = \mathrm{span}\{ 1, X_1, \dots, X_n \} \subset L_2$. Then

1. $\exists$ a unique $P_n X_{n+h} \in \mathcal{M}$ such that
$$\left\|X_{n+h} - P_n X_{n+h}\right\| = \inf_{Z \in \mathcal{M}} \left\|X_{n+h} - Z\right\|.$$
2. Residuals are orthogonal:
$$\tilde X_{n+h} = P_n X_{n+h} \iff \tilde X_{n+h} \in \mathcal{M},\ X_{n+h} - \tilde X_{n+h} \perp \mathcal{M},$$
i.e., for the inner product $\langle X, Y\rangle = \mathrm{E}(XY)$,
$$\langle X_{n+h} - \tilde X_{n+h}, 1\rangle = 0$$
$$\langle X_{n+h} - \tilde X_{n+h}, X_j\rangle = 0\ \forall j = 1, \dots n.$$

MSE Convergence (i.e., convergence in L^2)

$$
\begin{aligned}
X_n \xrightarrow{\text{MSE}} X &\iff \|X_n - X\|^2 = \mathrm{E}(X_n - X)^2 \longrightarrow 0 \text{ as } n \longrightarrow \infty \\
&\iff \|X_n - X_m\|^2 = \mathrm{E}(X_n - X_m)^2 \longrightarrow 0 \text{ as } n, m \longrightarrow \infty
\end{aligned}
$$

E.g., let $\{ X_t \}$ WS with mean μ and ACVF $\gamma(h)$ satisfying $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$,

$$
\bar{X}_n \xrightarrow{\text{MSE}} \mu, \quad \bar{X}_n \approx \mathcal{N}\left(\mu, \frac{1}{n} \sum_{h=-\infty}^{\infty} \gamma(h)\right).
$$

D-L Algorithm

Best linear predictor is

$$
P_n X_{n+1} = \sum_{j=1}^{n} \phi_{n,j} X_{n+1-j} = \boldsymbol{\gamma}'(1) \boldsymbol{\Gamma}_n^{-1} \begin{pmatrix} X_n \\ \vdots \\ X_1 \end{pmatrix}.
$$

Let $\{ X_t \}$ WS with mean 0 and ACVF $\gamma(h)$ satisfying $\gamma(0) > 0$, $\lim_{h \to \infty} \gamma(h) = 0$,

1. Set

$$
\begin{aligned}
P_0 X_1 &= 0, \\
\nu_0 &= \mathrm{E}(X_1 - P_0 X_1)^2 = \gamma(0).
\end{aligned}
$$

2. Set

$$
P_1 X_2 = \rho(1) X_1 = \frac{\gamma(1)}{\gamma(0)} X_1 = \phi_{1,1} X_1,
$$

$$
\nu_1 = \mathrm{E}(X_2 - P_1 X_2)^2 = \gamma(0) - \frac{[\gamma(1)]^2}{\gamma(0)} = \nu_0(1 - \phi_{1,1}^2).
$$

3. For $k \geq 2$, set

$$
\begin{aligned}
P_k X_{k+1} &= \sum_{j=1}^{k} \phi_{k,j} X_{k+1-j}, \\
\nu_k &= \mathrm{E}(X_{k+1} - P_k X_{k+1})^2 = (1 - \phi_{k,k}^2) \nu_{k-1},
\end{aligned}
$$

where

$$
\phi_{k,k} = \frac{\gamma(k) - \sum_{j=1}^{k-1} \phi_{k-1,j} \gamma(k-j)}{\nu_{k-1}},
$$

$$
\begin{pmatrix} \phi_{k,1} \\ \vdots \\ \phi_{k,k-1} \end{pmatrix} = \begin{pmatrix} \phi_{k-1,1} \\ \vdots \\ \phi_{k-1,k-1} \end{pmatrix} - \phi_{k,k} \begin{pmatrix} \phi_{k-1,k-1} \\ \vdots \\ \phi_{k-1,1} \end{pmatrix}.
$$

Partial Autocorrelation Function (PACF)

Correlation between X_1 and X_{n+1} after removing the effect of $X_2, \ldots, X_n$:

$$
\alpha(n) = \phi_{n,n} = \mathrm{Corr}(X_{n+1} - P_{\mathcal{K}_1} X_{n+1}, X_1 - P_{\mathcal{K}_1} X_1),
$$

where $\mathcal{K}_1 = \operatorname{span}\{ X_2, \ldots, X_n \}$.

Filtered Processes

Let $\{ Z_t \}$ be WS with mean zero and ACVF $\gamma_Z(h)$, and let $\psi_j \in \mathbb{R}\ \forall j \in \mathbb{Z}$ be absolutely summable (i.e., $\sum_{j=-\infty}^{\infty} \left|\psi_j\right| < \infty$). Then

$$
X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} = \left(\sum_{j=-\infty}^{\infty} \psi_j B^j\right) Z_t
$$

is WS with mean zero and ACVF

$$
\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Z(h - k + j).
$$

- Linear process: if $\{ Z_t \} \sim WN(0, \sigma^2)$, then

$$
\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \sigma^2 I(h - k + j = 0) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2.
$$

- MA(∞) process: if $\{ Z_t \} \sim WN(0, \sigma^2)$ and $\psi_j = 0\ \forall j < 0$, then

$$
X_t = \left(\sum_{j=0}^{\infty} \psi_j B^j\right) Z_t = \psi_0 Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \cdots,
$$

$$
\gamma_X(h) = \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \sigma^2.
$$

- AR(1) process: if $\{ Z_t \} \sim WN(0, \sigma^2)$ and $|\phi| < 1$, then $\sum_{j=0}^{\infty} \left|\phi^j\right| = \frac{1}{1-|\phi|} < \infty$ and

$$
X_t = \left(\sum_{j=0}^{\infty} \phi^j B^j\right) Z_t = \phi^0 Z_t + \phi^1 Z_{t-1} + \phi^2 Z_{t-2} + \cdots = Z_t + \phi X_{t-1},
$$

$$
\gamma_X(h) = \sum_{j=0}^{\infty} \phi^j \phi^{j+|h|} \sigma^2 = \frac{\sigma^2 \phi^{|h|}}{1 - \phi^2}.
$$

- AR(p) process: if $\{ Z_t \} \sim WN(0, \sigma^2)$ and Z_t uncorrelated with X_j $(j < t)$, then

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

has PACF

$$
\alpha(n) = \phi_{n,n} = \begin{cases} \phi n & \text{if } n = p \\ 0 & \text{if } n > p \end{cases}.
$$

ARMA Processes

- $\{ X_t \} \sim ARMA(p, q)$ for $p, q \in \mathbb{Z}_+$ if for any $t \in \mathbb{Z}$

$$
\phi(B) X_t = \theta(B) Z_t, \quad \{ Z_t \} \sim WN(0, \sigma^2),
$$

where the polynomials are

$$
\begin{aligned}
\text{AR} : \phi(z) &= 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p, \\
\text{MA} : \theta(z) &= 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q.
\end{aligned}
$$

- E.g., AR(1) = ARMA(1,0):

$$
\begin{aligned}
&X_t = \phi X_{t-1} + Z_t, \quad \{ Z_t \} \sim WN(0, \sigma^2). \\
&\text{AR} : \phi(z) = 1 - \phi z, \\
&\text{MA} : \theta(z) = 1.
\end{aligned}
$$

- E.g., MA(1) = ARMA(0,1):

$$
\begin{aligned}
&X_t = Z_t + \theta Z_{t-1}, \quad \{ Z_t \} \sim WN(0, \sigma^2). \\
&\text{AR} : \phi(z) = 1, \\
&\text{MA} : \theta(z) = 1 + \theta z.
\end{aligned}
$$

- $\{ X_t \}$ is WS $\iff$ $\phi(z)$ has no roots $z \in \mathbb{C}$ where $|z| = 1$.
- $\{ X_t \}$ is causal $\iff$ z is a root of $\phi(z) = 0$ then $|z| > 1$ $\iff$ $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, $\forall t \in \mathbb{Z}$, where $\sum_{j=1}^{\infty} \left|\psi_j\right| < \infty$ and

$$
\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}.
$$

- $\{ X_t \}$ is invertible $\iff$ z is a root of $\theta(z) = 0$ then $|z| > 1$ $\iff$ $Z_t = \sum_{j=0}^{\infty} \pi_j Z_{t-j}$, $\forall t \in \mathbb{Z}$, where $\sum_{j=1}^{\infty} \left|\pi_j\right| < \infty$ and

$$
\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}.
$$

- E.g., ARMA(1,1): $(1 - \phi B) X_t = (1 + \theta B) Z_t$, $\{ Z_t \} \sim WN(0, \sigma^2)$.
  Causal: $X_t = \frac{\theta(B)}{\phi(B)} Z_t = \psi(B) Z_t$, where $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$ and

$$
\begin{aligned}
z^0 &: \psi_0 = 1 \\
z^k &: \psi_k - \psi_{k-1} \phi = 0 \implies \psi_k = \phi^{k-1}(\theta + \phi), \ k \geq 1.
\end{aligned}
$$

  Then

$$
\gamma_X(h) = \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \sigma^2 = \sigma^2 \psi_{|h|} + \frac{(\theta + \phi)^2 \phi^{|h|}}{1 - \phi^2}.
$$

  Invertible: $Z_t = \frac{\phi(B)}{\theta(B)} X_t = \pi(B) X_t$, where

$$
\begin{aligned}
z^0 &: \pi_0 = 1 \\
z^k &: \pi_k + \pi_{k-1} \theta = 0 \implies \pi_k = (-\theta)^{k-1}(-\theta - \phi), \ k \geq 1.
\end{aligned}
$$

Yule-Walker Equations

Let $m = \max\{ p, q + 1 \}$, then

$$
\gamma(h) - \phi_1 \gamma(h-1) - \cdots - \phi_p \gamma(h-p) = \begin{cases} \sigma^2 \sum_{j=0}^{m} \theta_{j+h} \psi_j & \text{if } 0 \leq h < m \\ 0 & \text{if } h \geq m \end{cases},
$$

where $\theta_0 = 1$ and $\theta_j = 0$ for $j > p$.
For a pure AR(p) process, we have

$$
\begin{aligned}
h = 0 &\quad \gamma(0) - \phi_1 \gamma(1) - \cdots - \phi_p \gamma(p) = \sigma^2 \\
h = 1 &\quad \gamma(1) - \phi_1 \gamma(0) - \cdots - \phi_p \gamma(p-1) = 0 \\
&\vdots \\
h = p &\quad \gamma(p) - \phi_1 \gamma(p-1) - \cdots - \phi_p \gamma(0) = 0.
\end{aligned}
$$

ACVF of Causal ARMA(p,q)

Let $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}\ \forall t \in \mathbb{Z}$ where $\sum_{j=1}^{\infty} \left|\psi_j\right| < \infty$, $\{ X_t \} \sim WN(0, \sigma^2)$.

1. Direct: (need $\psi_0, \psi_1, \ldots$).

$$
\gamma_X(h) = \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \sigma^2.
$$

2. Explicit, non-recursive. Find $(\psi_0, \psi_1, \ldots, \psi_q)$. Compute k distinct roots $\xi_1, \ldots, \xi_k$ of $\phi(z) = 0$ (note $\{ X_t \}$ causal $\implies$ $\left|\xi_i\right| > 1$). Then $\gamma(h) = \sum_{i=1}^{k} \sum_{j=0}^{r_i - 1} B_{ij} \xi_i^{-h}$, for $h \geq m - p$, where $m = \max\{ p, q + 1 \}$ and r_i is the number of repeats of the i^{th} root. Find $\left\{ B_{ij} \right\}$, $\gamma(0), \ldots, \gamma(m - p - 1)$ by solving 1^{st} m Y-W equations:

$$
\begin{aligned}
h = 0 &\quad \gamma(0) - \phi_1 \gamma(1) - \cdots - \phi_p \gamma(p) = \sigma^2 \sum_{j=0}^{m} \theta_j \psi_j \\
h = 1 &\quad \gamma(1) - \phi_1 \gamma(0) - \cdots - \phi_p \gamma(p-1) = \sigma^2 \sum_{j=0}^{m} \theta_{j+1} \psi_j \\
&\vdots \\
h = m - 1 &\quad \gamma(m) - \phi_1 \gamma(m-1) - \cdots - \phi_p \gamma(m-p) \\
&\quad = \sigma^2 \sum_{j=0}^{m} \theta_{j+m-1} \psi_j
\end{aligned}
$$

3. Explicit, recursive. Find $(\psi_0, \psi_1, \ldots, \psi_q)$. Write down 1^{st} $p + 1$ Y-W equations

$$
\begin{aligned}
h = 0 &\quad \gamma(0) - \phi_1 \gamma(1) - \cdots - \phi_p \gamma(p) = a_1 \\
h = 1 &\quad \gamma(1) - \phi_1 \gamma(0) - \cdots - \phi_p \gamma(p-1) = a_2 \\
&\vdots \\
h = p &\quad \gamma(p) - \phi_1 \gamma(p-1) - \cdots - \phi_p \gamma(0) = a_{p+1}
\end{aligned}
$$

   Solve for the covariances, then $\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p)$ from Y-W equations for all $h \geq m$; solve for other covariances recursively.

Yule-Walker Estimation for AR(p)

1. Solve last p Y-W equations for $\phi_1, \ldots, \phi_p$: $\boldsymbol{\gamma}_p = \boldsymbol{\Gamma}_p \boldsymbol{\phi}_p =$

$$\begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}.$$

2. Return to 1st Y-W equation $\sigma^2 = \gamma(0) - \phi_1 \gamma(1) - \cdots - \phi_p \gamma(p)$.

3. Plug in $\hat{\gamma}(h)$ for $\gamma(h)$ (MME).

4. Calculate $\hat{\boldsymbol{\phi}}_{YW} = \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p$.

For AR(p),

$$\sqrt{n}(\hat{\boldsymbol{\phi}}_{YW} - \boldsymbol{\phi}_p) \xrightarrow{d} \mathcal{N}(\mathbf{0}_p, \sigma^2 \boldsymbol{\Gamma}_p^{-1})$$

For AR(1), $\hat{\phi}_{YW} \approx \mathcal{N}(\phi, (1-\phi^2)/n)$.
For AR(2), $\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} \approx \mathcal{N}\left(\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \frac{1}{n} \begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix} \right)$

Autocovariance Generating Function (ACGF)

If $\{X_t\}$ is WS with ACVF $\gamma(\cdot)$, then the ACGF of $\{X_t\}$ is

$$G(z) = \sum_{h=-\infty}^{\infty} \gamma(h) z^h,$$

provided this converges for all $z \in \mathbb{C}$ with $r^{-1} < |z| < 1$ for some $|r| > 1$.

- For $\{Z_t\} \sim WN(0, \sigma^2)$, $G(z) = \sum_{h=-\infty}^{\infty} \gamma(h) z^h = \gamma(0) z^0 = \sigma^2$.
- For a linear process with $\{Z_t\} \sim WN(0, \sigma^2)$ and $\psi_j \in \mathbb{R}\ \forall j \in \mathbb{Z}$ with $\sum_{j=-\infty}^{\infty} \left|\psi_j\right| < \infty$, the ACGF of $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} = \psi(B) Z_t$ is $G(z) = \sigma^2 \psi(z) \psi(z^{-1})$.
- For a filtered process $Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$, where $\{X_t\}$ is WS with ACGF $G_X(z)$ and $\psi_j \in \mathbb{R}\ \forall j \in \mathbb{Z}$ with $\sum_{j=-\infty}^{\infty} \left|\psi_j\right| < \infty$, the ACGF of $\{Y_t\}$ is $G_Y(z) = G_X(z)\psi(z)\psi(z^{-1})$.
- For a filtered process $Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$, where $\{X_t\}$ is WS with ACGF $G_X(z)$ and $\psi_j \in \mathbb{R}\ \forall j \in \mathbb{Z}$ with $\sum_{j=-\infty}^{\infty} \left|\psi_j\right| < \infty$, the ACGF of $\{Y_t\}$ is $G_Y(z) = G_X(z)\psi(z)\psi(z^{-1})$.
- For a WS ARMA $\{X_t\}$, we have $\phi(B)X_t = \theta(B)Z_t \implies X_t = \frac{\theta(B)}{\phi(B)} Z_t = \psi(B) Z_t$, where $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$. Then $G_X(z) = \sigma^2 \psi(z)\psi(z^{-1}) = \sigma^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}$.
- The sum $\{X_t\}$ of uncorrelated WS $\left\{X_{1,t}\right\}$ and $\left\{X_{2,t}\right\}$ has ACGF $G_X(z) = \sum_{h=-\infty}^{\infty} \gamma_X(h) z^h = \sum_{h=-\infty}^{\infty} [\gamma_{X_1}(h) + \gamma_{X_2}(h)] z^h = G_{X_1}(z) + G_{X_2}(z)$.

Complex Numbers

For $i = \sqrt{1}$ and $\omega \in (-\pi, \pi]$ define sinusoid with frequency ω:

$$e^{it\omega} = \cos(t\omega) + i \sin(t\omega).$$

For complex $z = a + ib \in \mathbb{C}$ define complex conjugate $\overline{z} = a - ib$ and modulus $|z| = \sqrt{a^2 + b^2} \implies |z|^2 = z\overline{z}$.
Since cosine is even and sin is odd, $\overline{e^{it\omega}} = e^{-it\omega}$.
Inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i \overline{y_i}$ defined for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.

Fourier Frequencies

Defined as $\omega_j = 2\pi j/n$ for all $j \in \mathcal{F}_n$ where

$$\mathcal{F}_n = \{ -\lfloor (n-1)/2 \rfloor, \ldots, -1, 0, 1, \ldots, \lfloor n/2 \rfloor \}.$$

For $j \in \mathcal{F}_n$, we have $\omega_j \in (-\pi, \pi]$.
For odd n, we have n frequencies:

$$\mathcal{F}_n = \{ -(n-1)/2, \ldots, -1, 0, 1, \ldots, (n-1)/2 \}.$$

For even n, we have n frequencies ($\omega_{n/2} = \pi$):

$$\mathcal{F}_n = \{ -(n-2)/2, \ldots, -1, 0, 1, \ldots, n/2 \}.$$

$\left\{ \mathbf{e}_j : j \in \mathcal{F}_n \right\}$ is an orthonormal basis for $\mathbb{C}^n$, where

$$\mathbf{e}_j = \frac{1}{\sqrt{n}} \left(e^{i\omega_j}, e^{i2\omega_j}, \ldots, e^{in\omega_j} \right) \in \mathbb{C}^n,$$

meaning $\forall \mathbf{y} \in \mathbb{C}^n, \exists a_j$ s.t. $\mathbf{y} = \sum_{j \in \mathcal{F}_n} a_j \mathbf{e}_j$ and $\langle \mathbf{e}_j, \mathbf{e}_k \rangle = I(j = k \in \mathcal{F}_n)$.

Discrete Fourier Transform

For $\mathbf{X} = (X_1, \ldots, X_n) \in \mathbb{R}^n$ it holds that

$$\mathbf{X} = \sum_{j \in \mathcal{F}_n} d_j X_j,$$

where $d_j = \langle \mathbf{X}, \mathbf{e}_j \rangle \in \mathbb{C}$, $j \in \mathcal{F}_n$. Then $\left\{ d_j = \langle \mathbf{X}, \mathbf{e}_j \rangle : j \in \mathcal{F}_n \right\}$ is the discrete Fourier transform of $\mathbf{X}$.

Periodogram

Periodogram of $\mathbf{X}$ at frequency ω_j is

$$I_n(\omega_j) = d_j \overline{d_j} = \left| d_j \right|^2 = \left| \langle \mathbf{X}, \mathbf{e}_j \rangle \right|^2 = \frac{1}{n} \left| \sum_{t=1}^{n} X_t e^{-it\omega_j} \right|^2.$$

Properties:

- At $j = 0 \implies \omega_j = \omega_0 = 0 \implies \mathbf{e}_0 = (1, \ldots, 1)'/\sqrt{n}$,

$$I_n(0) = \frac{1}{n} \left| \sum_{t=1}^{n} X_t \cdot 1 \right|^2 = n(\bar{X}_n)^2.$$

- $d_j, I_n(\omega_j)$ are not affected by sample mean corrections with frequencies $\omega_j \neq 0$:

$$\langle \mathbf{X} - \bar{X}_n \sqrt{n} \mathbf{e}_0, \mathbf{e}_j \rangle = d_j - \bar{X}_n \sqrt{n} I(j = 0).$$

- Symmetric: $I_n(\omega_j) = I_n(-\omega_j) = I_n(\omega_{-j})$.
- Sum of squares total property:

$$\sum_{t=1}^{n} X_t^2 = \langle \mathbf{X}, \mathbf{X} \rangle = \sum_{j \in \mathcal{F}_n} I_n(\omega_j).$$

- Related to sample ACVF:

$$I_n(\omega_j) = \sum_{k=-(n-1)}^{n-1} \hat{\gamma}(k) e^{-ik\omega_j}.$$

Spectral Density

The spectral density of WS $\{X_t\}$ with ACVF $\gamma(\cdot)$ is

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\omega} = \frac{1}{2\pi} G(e^{-i\omega}), \quad \omega \in [-\pi, \pi].$$

Properties:

- Nonnegative: $f(\omega) \geq 0$, $\omega \in [-\pi, \pi]$ (by NND of γ).
- Symmetric: $f(-\omega) = f(\omega)$.
- For any $k \in \mathbb{Z}$, $\int_{-\pi}^{\pi} e^{ik\omega} f(\omega) d\omega = \gamma(k)$.

Obtaining Spectral Densities

- WN: spectral density f of $\{Z_t\} \sim WN(0, \sigma^2)$ is

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\omega} = \frac{1}{2\pi} \gamma(0) e^{-i0\omega} = \frac{\sigma^2}{2\pi}, \quad \omega \in [-\pi, \pi].$$

- Filtered process: let $\{X_t\}$ be WS with ACGF $G_X(\cdot)$ and spectral density $f_X(\cdot)$. For $\psi_j \in \mathbb{R}\ \forall j \in \mathbb{Z}$ with $\sum_{j=-\infty}^{\infty} \left|\psi_j\right| < \infty$, define $Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} = \psi(B) X_t$. Then the spectral density of $\{Y_t\}$ is

$$f_Y(\omega) = f_X(\omega) \left| \psi(e^{i\omega}) \right|^2, \quad \omega \in [-\pi, \pi].$$

- WS ARMA: $\phi(B)X_t = \theta(B)Z_t$, $\{Z_t\} \sim WN(0, \sigma^2)$ where $\theta(z) \neq 0$ for $|z| = 1$.

$$f_X(\omega) = \frac{\left|\theta(e^{i\omega})\right|^2}{\left|\phi(e^{i\omega})\right|^2} \frac{\sigma^2}{2\pi}, \quad \omega \in [-\pi, \pi].$$

- AR(2): $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$

$$f_X(\omega) = \frac{\sigma^2}{2\pi \left[1 + \phi_1^2 + \phi_2^2 - 2\phi_1 \cos(\omega) - 2\phi_2 \cos(2\omega) + 2\phi_1\phi_2 \cos(\omega) \right]}.$$

Power Transfer Function (PTF)

For WS $\{X_t\}$ and $\psi_j \in \mathbb{R}\ \forall j \in \mathbb{Z}$ with $\sum_{j=-\infty}^{\infty} \left|\psi_j\right| < \infty$, define the WS filtered process $Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} = \psi(B) X_t$. Then the PTF of $\{Y_t\}$ is

$$\left| \phi(e^{i\omega}) \right|^2, \ \omega \in [-\pi, \pi].$$

E.g., lag-k difference filter $\psi(B) = 1 - B^k$ has PTF $\left|\psi(e^{i\omega})\right|^2 = \left|1 - e^{i\omega k}\right|^2 = 2 - 2\cos(\omega k)$. Then PTF is zero for $\omega = 2\pi p/k$ (for integer p).

Distributional Properties of Periodogram

Let $\{X_t\}$ be WS with ACVF $\gamma(\cdot)$, spectral density $f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\omega}$, $\omega \in [-\pi, \pi]$, and periodogram

$$I_n(\omega_j) = \begin{cases} n(\bar{X}_n)^2 & \text{if } \omega_j = 0 (j = 0) \\ \sum_{h=-(n-1)}^{(n-1)} \hat{\gamma}(h) e^{-ih\omega_j} & \text{if } \omega_j \neq 0 \in \mathcal{F}_n \end{cases}$$

- For $\omega = 0$ and $\mathrm{E}(X_t) = \mu$,

$$\lim_{n \to \infty} \mathrm{E}\left[I_n(0) - n\mu^2 \right] = 2\pi f(0) = \sum_{k=-\infty}^{\infty} \gamma(k).$$

Hence for large n, $\mathrm{E}[I_n(0)] \approx n\mu^2 + 2\pi f(0)$. In the frequency domain,

$$\bar{X}_n \approx \mathcal{N}\left(\mu, \frac{2\pi f(0)}{n} \right).$$

For WS ARMA(p, q) process $\{X_t\}$ with $\phi(x) \neq 0$ for $|z| = 1$,

$$\bar{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n} \frac{(1 + \theta_1 + \cdots + \theta_q)^2}{1 - \phi_1 - \cdots - \phi_p)^2} \right).$$

- For $\omega \in [-\pi, \pi] \setminus \{0\}$ (i.e., not just non-zero Fourier freq), it holds that

$$\lim_{n \to \infty} \mathrm{E}\left[\frac{I_n(\omega)}{2\pi} \right] = f(\omega).$$

Hence for large n, $\frac{I_n(\omega)}{2\pi}$ is an unbiased estimator for $f(\omega)$.

- For $\{ Z_t \} \overset{\text{iid}}{\sim} (0, \sigma^2)$ and $\psi_j \in \mathbb{R}\ \forall j \in \mathbb{Z}$ with $\sum_{j=-\infty}^{\infty} \left|\psi_j\right| < \infty$, define the linear process $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ with spectral density f. If $f(\omega) > 0$ for all $\omega \in [-\pi, \pi]$ and $0 < \lambda_1 < \cdots < \lambda_m < \pi$ are a set of *fixed* frequencies, then
$$\frac{1}{2\pi} I_n(\lambda_i) \approx \text{ind Exp}(f(\lambda_i)).$$
So
$$\mathrm{E}\left[\frac{1}{2\pi} I_n(\lambda_i)\right] \approx f(\lambda_i), \quad \mathrm{Var}\left[\frac{1}{2\pi} I_n(\lambda_i)\right] \approx [f(\lambda_i)]^2,$$
and since $Y \sim \text{Exp}(\theta) \implies 2Y/\theta \sim \text{Exp}(2) \sim \chi^2_2$,
$$\frac{1}{2\pi} I_n(\lambda_i) \frac{2}{f(\lambda_i)} \approx \chi^2_2.$$

- If $\{ X_t \} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2$, then the above results are exact for all n and all Fourier frequencies except $j = 0, n/2$:
$$\left\{ \frac{I_n(\omega)}{2\pi} : \omega_j \in \mathcal{F}_n, \omega_j \notin \{ 0, \pi \} \right\}.$$

- Periodogram is not a consistent estimator of the spectral density: for $\lambda \in (0, \pi)$,
$$\frac{I_n(\lambda)}{2\pi} \xrightarrow{d} \text{Exp}[f(\lambda)]$$
but
$$\frac{I_n(\lambda)}{2\pi} \overset{p}{\not\to} f(\lambda).$$
Instead use window estimator of periodogram,
$$\hat{f}(\lambda) = \sum_{|k| < m_n} W_n(k) \frac{I_n(\omega_{j+k})}{2\pi},$$
where $W_n(\cdot)$ is a weight function, m_n is a bandwidth, and $\omega_j \in \mathcal{F}_n$ is closest to λ. Under some conditions, $\hat{f}(\lambda)$ will be MSE-consistent for $f(\lambda)$: as $n \to \infty$,
$$\mathrm{E}[\hat{f}(\lambda)] \to f(\lambda),$$
$$\frac{\mathrm{Cov}[\hat{f}(\lambda), \hat{f}(\omega)]}{\sum_{|k| \le m_n} [W_n(k)]^2} \to \begin{cases} 2[f(\lambda)]^2 & \text{if } \lambda = \omega \in \{ 0, \pi \} \\ 1[f(\lambda)]^2 & \text{if } \lambda = \omega \notin \{ 0, \pi \} \\ 0 & \text{if } \lambda \neq \omega \end{cases}$$

Least Squares Estimation for AR(p)

From $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$, given data $X_1, \dots, X_n$, use multiple regression to
$$\text{regress} \begin{pmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_n \end{pmatrix} \text{ on } \begin{pmatrix} X_p & X_{p-1} & \cdots & X_1 \\ X_{p+1} & X_p & \cdots & X_2 \\ \vdots & \vdots & \vdots & \vdots \\ X_{n-1} & X_{n-2} & \cdots & X_{n-p} \end{pmatrix}$$

Maximum Likelihood for ARMA

Model parameters are $\boldsymbol{\Psi} = (\sigma^2, \boldsymbol{\phi}, \boldsymbol{\theta})$. Then the likelihood for the data $\mathbf{X}_n$ is, by the "chain rule",
$$L(\boldsymbol{\Psi} \mid \mathbf{X}_n) = P_{\boldsymbol{\Psi}}(X_n \mid \mathbf{X}_{n-1}) \dots P_{\boldsymbol{\Psi}}(X_2 \mid X_1) P_{\boldsymbol{\Psi}}(X_1).$$
For Gaussian WS $\{ X_t \}$ with mean zero,
$$\mathbf{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \boldsymbol{\Gamma}_n = [\gamma(i-j)]_{i,j=1,\dots,n} \right],$$
where the parameters $\boldsymbol{\Psi}$ appear in $\gamma(\cdot)$. Then
$$P_{\boldsymbol{\Psi}}(X_t \mid \mathbf{X}_{t-1}) \sim \mathcal{N}\left(\mathrm{E}(X_t \mid \mathbf{X}_{t-1}) = \hat{X}_t, \mathrm{Var}(X_t \mid \mathbf{X}_{t-1}) \right),$$
where $\hat{X}_t = P_{t-1} X_t = \phi_{t-1,1} X_t + \dots \phi_{t-1,t-1} X_1$ is the best linear predictor of X_t given $(X_1, \dots, X_{t-1}$ and the coefficients
$$\begin{pmatrix} \phi_{t-1,1} \\ \vdots \\ \phi_{t-1,t-1} \end{pmatrix} = \boldsymbol{\Gamma}_n^{-1} \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(t-1) \end{pmatrix}$$
depend on ARMA parameters $\boldsymbol{\phi}, \boldsymbol{\theta}$ through the ACVF $\gamma(\cdot)$ and can be obtained via the D-L algo. So,
$$P_{\boldsymbol{\Psi}}(X_t \mid \mathbf{X}_{t-1}) = (2\pi)^{-1/2} (\sigma^2 r_{t-1})^{-1/2} \exp\left[-(X_t - \hat{X}_t)^2 / (2\sigma^2 r_{t-1}) \right],$$
where $\sigma^2 r_{t-1} = \mathrm{E}(X_t - \hat{X}_t)^2 = \mathrm{Var}(X_t \mid \mathbf{X}_{t-1})$.
Profile out σ^2 to get MLE of σ^2 given $\hat{\boldsymbol{\phi}}_{\text{MLE}}, \hat{\boldsymbol{\theta}}_{\text{MLE}}$:
$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n} \sum_{t=1}^{n} \frac{(X_t - \hat{X}_t)^2}{r_{t-1}}.$$

Distribution of ARMA MLE

ARMA process $\phi(B) X_t = \theta(B) Z_t$, $\{ Z_t \} \sim WN(0, \sigma^2)$. Define new WN process $Z_t^* \sim WN(0, \sigma^2)$ and new processes $U_t = Z_t^* / \phi(B) \sim \text{AR}(p)$ and $V_t = Z_t^* / \theta(B) \sim \text{AR}(q)$. Then
$$\begin{pmatrix} \hat{\boldsymbol{\phi}}_{\text{MLE}} \\ \hat{\boldsymbol{\theta}}_{\text{MLE}} \end{pmatrix} \approx \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\phi}_{\text{MLE}} \\ \boldsymbol{\theta}_{\text{MLE}} \end{pmatrix}, \frac{\sigma^2}{n} \begin{pmatrix} \mathrm{E}(\mathbf{U}_p \mathbf{U}_p') & \mathrm{E}(\mathbf{U}_p \mathbf{V}_p') \\ \mathrm{E}(\mathbf{V}_q \mathbf{U}_p') & \mathrm{E}(\mathbf{V}_q \mathbf{V}_q') \end{pmatrix} \right),$$
where $\mathbf{U}_p = (U_p, \dots, U_1)'$ and $\mathbf{V}_q = (V_q, \dots, V_1)'$.

State Space Model

$$\underset{(w \times 1)}{\mathbf{Y}_t} = \underset{(w \times v)}{\mathbf{G}_t} \underset{(v \times 1)}{\mathbf{X}_t} + \underset{(w \times 1)}{\mathbf{W}_t} \qquad \text{(Observation Equation)}$$
$$\underset{(v \times 1)}{\mathbf{X}_{t+1}} = \underset{(v \times v)}{\mathbf{F}_t} \underset{(v \times 1)}{\mathbf{X}_t} + \underset{(v \times 1)}{\mathbf{V}_t} \qquad \text{(State Equation)}$$
where

- $\left\{ \begin{pmatrix} \mathbf{V}_t \\ \mathbf{W}_t \end{pmatrix} \right\}_{t \ge 1}$ are uncorrelated random vectors,
- $\mathrm{E}(\mathbf{W}_t) = \mathbf{0}_w$, $\mathrm{E}(\mathbf{V}_t) = \mathbf{0}_v$,
- $\mathrm{Var}\begin{pmatrix} \mathbf{V}_t \\ \mathbf{W}_t \end{pmatrix} = \begin{pmatrix} \mathrm{E}(\mathbf{V}_t \mathbf{V}_t') & \mathrm{E}(\mathbf{V}_t \mathbf{W}_t') \\ \mathrm{E}(\mathbf{W}_t \mathbf{V}_t') & \mathrm{E}(\mathbf{W}_t \mathbf{W}_t') \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_t & \mathbf{S}_t \\ \mathbf{S}_t' & \mathbf{R}_t \end{pmatrix}$,
- For each t, $\mathbf{V}_t$ and $\mathbf{W}_t$ are uncorrelated with $\{ \mathbf{X}_s : s \le t \}$.

Note: WS $\implies$ time-invariant (but not vice versa).
E.g., Random Walk + Noise
$$Y_t = X_t + W_t, \quad X_{t+1} = X_t + V_t,$$
for $\{ W_t \} \sim WN(0, \sigma_w^2), \{ V_t \} \sim WN(0, \sigma_v^2)$ are uncorrelated. Here, $F_t = G_t = 1, Q_t = \sigma_v^2, R_t = \sigma_\omega^2, S_t = 0$.
E.g., linear model $Y_t = \mathbf{Z}_t' \boldsymbol{\beta} + W_t$, $\{ X_t \} \sim WN(0, \sigma^2)$. Then $\mathbf{F}_t = I_{v \times v}, \mathbf{V}_t = (0, \dots, 0)'$, and
$$Y_t = \mathbf{Z}_t' \boldsymbol{\beta} + W_t \qquad \text{(Observation Equation)}$$
$$\mathbf{X}_{t+1} = \boldsymbol{\beta} \qquad \text{(State Equation)}$$

ARMA Models in State Space Form

AR(p): $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$, $\{ Z_t \} \sim WN(0, \sigma^2)$.
$$\underset{(p \times 1) \mathbf{X}_{t+1}}{\begin{pmatrix} X_{t+1} \\ X_t \\ \vdots \\ X_{t+2-p} \end{pmatrix}} = \underset{(p \times p) \mathbf{F}_t}{\begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \cdots & 1 & 0 \end{pmatrix}} \underset{(p \times 1) \mathbf{X}_t}{\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t+1-p} \end{pmatrix}} + \underset{(p \times 1) \mathbf{V}_t}{\begin{pmatrix} Z_{t+1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}} \qquad \text{(State)}$$
$$Y_t = \underset{(1 \times p) \mathbf{G}_t}{(1 \quad 0 \quad \dots \quad 0)} \mathbf{X}_t + \underset{(1 \times 1) \mathbf{W}_t}{0} = X_t \qquad \text{(Obs)}$$

MA(1): $X_t = Z_t + \theta Z_{t-1}$, $\{ Z_t \} \sim WN(0, \sigma^2)$. Let $\underset{(2 \times 1)}{\mathbf{X}_t} = \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}$, then
$$\mathbf{X}_{t+1} = \underset{(2 \times 2) \mathbf{F}_t}{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}} \mathbf{X}_t + \underset{(2 \times 1) \mathbf{V}_t}{\begin{pmatrix} Z_{t+1} \\ \theta Z_{t+1} \end{pmatrix}} \qquad \text{(State)}$$
$$Y_t = \underset{(1 \times 2) \mathbf{G}_t}{(1 \quad 0)} \mathbf{X}_t + \underset{(1 \times 1) \mathbf{W}_t}{0} = X_{1,t} \qquad \text{(Obs)}$$

Co-integration

Real-valued $\{ Y_t \} \sim I(d)$ if $\left\{ (1-B)^d Y_t \right\}$ is WS but $\left\{ (1-B)^{d-1} Y_t \right\}$ is not WS.
Random vectors $\{ \mathbf{Y}_t \} \sim I(d)$ if each component is $I(d)$.
$\{ \mathbf{Y}_t \} \sim I(d)$ is co-integrated with co-integrating factor $\boldsymbol{\alpha}$ is $\left\{ \boldsymbol{\alpha}' \mathbf{Y}) t \right\} \sim I(k)$ for some $k < d$.
E.g., drunk guy and puppy example where $\{ Y \}_t \sim I(d = 1)$:
$$\mathbf{Y}_t = \begin{pmatrix} d_t \\ p_t \end{pmatrix} = \begin{pmatrix} 1 \\ \gamma \end{pmatrix} d_t + \begin{pmatrix} 0 \\ w_t \end{pmatrix} \qquad \text{(Observation Equation)}$$
$$d_{t+1} = d_t + v_t \qquad \text{(State Equation)}$$
for $\{ W_t \} \sim WN(0, \sigma_w^2$, $\{ V_t \} \sim WN(0, \sigma_v^2$. Then $\mathbf{Y}_t$ is co-integrated with factor $\boldsymbol{\alpha} = \begin{pmatrix} \gamma \\ -1 \end{pmatrix}$ because $\boldsymbol{\alpha}' \mathbf{Y}_t = -W_t \sim I(d = 0)$ is WS.

Kalman Filter

$\mathbf{X}_{t|k}$ predicts $\mathbf{X}_t$ based on past observations (not states) $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_k$ with error covariance matrix for $\mathbf{X}_{t|k}$ given by
$$\boldsymbol{\Omega}_{t|k} = \mathrm{E}\left[(\mathbf{X}_t - \mathbf{X}_{t|k})(\mathbf{X}_t - \mathbf{X}_{t|k})' \right].$$
Goals:

1. One-step ahead prediction (predict $\mathbf{X}_t$ from $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$):
$$\hat{\mathbf{X}}_t \equiv \mathbf{X}_{t|t-1}, \quad \boldsymbol{\Omega}_t \equiv \boldsymbol{\Omega}_{t|t-1}.$$
2. Filtering (predict $\mathbf{X}_t$ from $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_t$):
$$\mathbf{X}_{t|t}, \quad \boldsymbol{\Omega}_{t|t}.$$
3. Smoothing (predict $\mathbf{X}_t$ from $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_n$):
$$\mathbf{X}_{t|n}, \quad \boldsymbol{\Omega}_{t|n}, \quad t \le n.$$

Kalman Filter Steps

Assume that $\mathbf{S}_t = \mathbf{0}_{w \times v}$ in Kalman Filter, $\mathbf{F}_t, \mathbf{G}_t, \mathbf{Q}_t, \mathbf{R}_t$ are known.

1. Start-up: $\hat{X}_{1|0} = \mathbf{X}_{1|0}$ (often $E(\mathbf{X}_1)$), and $\boldsymbol{\Omega}_1 = \mathrm{E}\left[(\mathbf{X}_1 - \hat{\mathbf{X}}_1)(\mathbf{X}_1 - \hat{\mathbf{X}}_1)' \right]$.
2. Innovation at $t \ge 1$: (new $\mathbf{Y}_t$ becomes available in addition to $\mathbf{Y}_0, \dots, \mathbf{Y}_{t-1}$)
$$\mathbf{I}_t = \mathbf{Y}_t - \hat{\mathbf{Y}}_t = \mathbf{Y}_t - \mathbf{G}_t \hat{\mathbf{X}}_t = \mathbf{G}(\mathbf{X}_t - \hat{\mathbf{X}}_t) + \mathbf{W}_t.$$
$$\boldsymbol{\Delta}_t = \mathrm{Var}(\mathbf{I}_t) = \mathrm{E}(\mathbf{I}_t \mathbf{I}_t') = \mathbf{G}_t \boldsymbol{\Omega}_t \mathbf{G}_t' + \mathbf{R}_t \quad (*)$$
3. Filter (update) at $t \ge 1$:
$$\mathbf{X}_{t|t} = \hat{\mathbf{X}}_t + \boldsymbol{\Omega}_t \mathbf{G}_t' \boldsymbol{\Delta}_t^{-1} \mathbf{I}_t$$
$$\boldsymbol{\Omega}_{t|t} = \boldsymbol{\Omega}_t - \boldsymbol{\Omega}_t \mathbf{G}_t' \boldsymbol{\Delta}_t^{-1} \mathbf{G}_t \boldsymbol{\Omega}_t \quad (*)$$
Note: $\mathbf{X}_{t|t}$ is $\mathrm{E}(\mathbf{X}_t | \mathbf{Y}_0, \dots, \mathbf{Y}_t)$ assuming Gaussian processes or best linear predictor.
4. Predict at $t \ge 1$: (prediction of $\mathbf{X}_{t+1}$ from $\mathbf{Y}_0, \dots, \mathbf{Y}_t$)
$$\hat{\mathbf{X}}_{t+1} = \mathbf{F}_t \mathbf{X}_{t|t}$$
$$\boldsymbol{\Omega}_{t+1} = \mathrm{E}\left[(\mathbf{X}_{t+1} - \hat{\mathbf{X}}_{t+t})(\mathbf{X}_{t+1} - \hat{\mathbf{X}}_{t+t})' \right] = \mathbf{F}_t \boldsymbol{\Omega}_{t|t} \mathbf{F}_t' + \mathbf{Q}_t \quad (*)$$