

MATH 545 Winter 2025: Final Formula Sheet

- **Multivariate Normal distribution results:**

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

with $\mathbf{X}_1$ and $\mathbf{X}_2$ being $k_1 \times 1$ and $k_2 \times 1$ respectively, then

$$\mathbf{X}_1 \sim \mathcal{N}_{k_1}(\boldsymbol{\mu}_1, \Sigma_{11})$$

$$\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1 \sim \mathcal{N}_{k_2}(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}).$$

In the time series context, if we consider the one-step-ahead (conditional) probability distribution

$$X_{n+1} | X_{1:n} = x_{1:n}$$

it follows by standard properties that

$$X_{n+1} | X_{1:n} = x_{1:n} \sim \mathcal{N}(\mu_{(n+1)|n}, \sigma_{(n+1)|n}^2).$$

If $\gamma_h = \gamma_X(h)$ for each h is the autocovariance function (ACVF), the covariance matrix of $X_{1:(n+1)}$ is

$$\mathbf{\Gamma}_X(n+1) = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-2} & \gamma_{n-1} & \gamma_n \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-3} & \gamma_{n-2} & \gamma_{n-1} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-4} & \gamma_{n-3} & \gamma_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{n-2} & \gamma_{n-3} & \gamma_{n-4} & \cdots & \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_n & \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix} = \begin{bmatrix} \mathbf{\Gamma}_X(n) & \boldsymbol{\gamma}_n^R \\ \{\boldsymbol{\gamma}_n^R\}^\top & \gamma_0 \end{bmatrix}$$

where $\boldsymbol{\gamma}_n^R = [\gamma_n, \gamma_{n-1}, \gamma_{n-1}, \dots, \gamma_2, \gamma_1]^\top$. Then

$$\mu_{(n+1)|n} = \mu_X + \{\boldsymbol{\gamma}_n^R\}^\top \{\mathbf{\Gamma}_X(n)\}^{-1} (x_{1:n} - \mu_X \mathbf{1}_n)$$

$$\sigma_{(n+1)|n}^2 = \gamma_0 - \{\boldsymbol{\gamma}_n^R\}^\top \{\mathbf{\Gamma}_X(n)\}^{-1} \boldsymbol{\gamma}_n^R.$$

- **Linear processes:** $\{X_t\}$ is a linear process if for all t

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

where $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$ and $\{\psi_j\}$ is a real sequence with

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

We have that

$$\gamma_X(h) = \mathbb{E}[X_t X_{t+h}] = \sigma_Z^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}$$

- **Optimal linear forecasting:** The minimum Mean-Square Error (MSE) criterion for forecasting X_{n+h} from $X_1, \dots, X_n$ is based on choosing constants $a_{0:n} = (a_0, a_1, \dots, a_n)^\top$ to minimize

$$\mathbb{E} \left[\left(X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n-i+1} \right)^2 \right]$$

where the expectation is over the joint distribution of all the random quantities. The linear predictor

$$P_n X_{n+h} = a_0 + \sum_{i=1}^n a_i X_{n-i+1}$$

solves simultaneously

$$a_0 = \mu \left(1 - \sum_{i=1}^n a_i \right) \quad \mathbf{\Gamma}_n \mathbf{a} = \boldsymbol{\gamma}_{h:(h+n-1)}$$

where

$$\mathbf{\Gamma}_n = [\gamma_X(i-j)]_{ij} \quad \mathbf{a} = (a_1, \dots, a_n)^\top$$

and $\boldsymbol{\gamma}_{h:(h+n-1)} = (\gamma_X(h), \dots, \gamma_X(h+n-1))^\top$. The solution $\mathbf{a}(n, h) = (a(1, h), \dots, a(n, h))^\top$ yields

$$P_n X_{n+h} = \mu + \sum_{i=1}^n a(i, h)(X_{n-i+1} - \mu).$$

By construction $\mathbb{E}[(X_{n+h} - P_n X_{n+h})] = 0$ and

$$\mathbb{E}[(X_{n+h} - P_n X_{n+h})X_j] = 0 \quad j = 1, \dots, n$$

The minimum value of the MSE achieved by $\mathbf{a}(n, h)$ is

$$\mathbb{E}[(X_{n+h} - P_n X_{n+h})^2] = \gamma_X(0) - \mathbf{a}(n, h)^\top \boldsymbol{\gamma}_{h:(h+n-1)}$$

- **ARMA processes:** $\{X_t\}$ is an ARMA(p, q) process if it is stationary and satisfies

$$\Phi(B)X_t = \Theta(B)Z_t$$

where $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p = \prod_{j=1}^p (1 - \xi_j z) \quad \Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q = \prod_{j=1}^q (1 - \omega_j z).$$

The model is stationary and causal if and only if $|\xi_j| < 1$ for $j = 1, \dots, p$. In this case, the linear process representation of $\{X_t\}$ is defined by

$$\Psi(z) = \frac{\Theta(z)}{\Phi(z)}.$$

The ACVF for an ARMA(p, q) process can be found via the infinite system of equations

$$\gamma_X(h) - \sum_{j=1}^p \phi_j \gamma_X(h-j) = \begin{cases} \sigma_Z^2 \sum_{j=0}^m \theta_{j+h} \psi_j & 0 \leq h < m \\ 0 & h \geq m \end{cases} \quad m = \max\{p, q+1\}.$$

The autocovariance generating function (ACVGF) is given by

$$G_X(z) = \sum_{h=-\infty}^{\infty} \gamma_X(h) z^h = \sigma_Z^2 \Psi(z) \Psi(z^{-1}) = \sigma_Z^2 \frac{\Theta(z) \Theta(z^{-1})}{\Phi(z) \Phi(z^{-1})} \quad 1/r < |z| < r.$$

- **Non-stationary models:** If d is a non-negative integer, then $\{X_t\}$ is an ARIMA(p, d, q) process if $\{Y_t\}$ defined by

$$Y_t = (1 - B)^d X_t$$

is a (causal) ARMA(p, q) process, that is

$$\Phi(B)(1 - B)^d X_t = \Theta(B)Z_t$$

where $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$; $\{X_t\}$ is stationary only if $d = 0$.

If d and D are non-negative integers, then $\{X_t\}$ is a SARIMA process with period s

$$\{X_t\} \sim \text{ARIMA}(p, d, q) \times (P, D, Q)_s$$

if

$$Y_t = (1 - B)^d (1 - B^s)^D X_t$$

is a causal ARMA process determined by

$$\Phi(B)\Phi_S(B^s)Y_t = \Theta(B)\Theta_S(B^s)Z_t$$

where $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$ and

$$\begin{aligned} \Phi(z) &= 1 - \phi_1 z - \dots - \phi_p z^p & \Theta(z) &= 1 + \theta_1 z + \dots + \theta_q z^q \\ \Phi_S(z) &= 1 - \phi_1^{(S)} z - \dots - \phi_P^{(S)} z^P & \Theta_S(z) &= 1 + \theta_1^{(S)} z + \dots + \theta_Q^{(S)} z^Q \end{aligned}$$

- **Vector and state space models:** $\mathbf{X}_t = (X_{1,t}, \dots, X_{K,t})^\top$ is weakly stationary if

$$\mathbb{E}[\mathbf{X}_t] = \boldsymbol{\mu} \quad \text{Cov}[\mathbf{X}_t, \mathbf{X}_{t-h}] = \mathbb{E}[(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_{t-h} - \boldsymbol{\mu})^\top] = \boldsymbol{\Gamma}_h$$

and $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}_h$ are a constant $(K \times 1)$ vector, and $(K \times K)$ matrix that depends on h but not t , respectively. If $\mathbf{D} = \text{diag}(\sqrt{\Gamma_{11}(0)}, \dots, \sqrt{\Gamma_{KK}(0)})$, the lag-zero cross-correlation matrix is

$$\boldsymbol{\rho}_0 = \mathbf{D}^{-1} \boldsymbol{\Gamma}_0 \mathbf{D}^{-1}$$

with (i, j) th element

$$\rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}}$$

The lag- h cross-correlation matrix is $\boldsymbol{\rho}_h = \mathbf{D}^{-1} \boldsymbol{\Gamma}_h \mathbf{D}^{-1}$. In general $\Gamma_{ij}(h) = \Gamma_{ji}(-h)$, so $\boldsymbol{\Gamma}_h = \boldsymbol{\Gamma}_{-h}^\top$.

The zero-mean VARMA(p, q) model

$$\Phi(B)\mathbf{X}_t = \Theta(B)\mathbf{Z}_t$$

where $\mathbf{Z}_t \sim \text{WN}(\mathbf{0}, \Sigma)$ and

$$\Phi(B) = \mathbf{I}_K - \Phi_1 B - \dots - \Phi_p B^p \quad \Theta(B) = \mathbf{I}_K + \Theta_1 B + \dots + \Theta_q B^q$$

has the linear process representation

$$\mathbf{X}_t = \{\Phi(B)\}^{-1} \Theta(B) \mathbf{Z}_t = \Psi(B) \mathbf{Z}_t$$

and the ACVF matrix sequence can be computed for $h \geq 0$ using

$$\boldsymbol{\Gamma}_h = \mathbb{E}[\mathbf{X}_t \mathbf{X}_{t-h}^\top] = \sum_{j=0}^{\infty} \Psi_{j+h} \Sigma (\Psi_j)^\top.$$

Key recursive computations for state space models are

$$\begin{aligned}
\text{Prediction:} \quad p(\mathbf{x}_t | \mathbf{y}_{1:(t-1)}) &= \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{y}_{1:(t-1)}) d\mathbf{x}_{t-1} \\
\text{Filtering:} \quad p(\mathbf{x}_t | \mathbf{y}_{1:t}) &= \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:(t-1)})}{p(\mathbf{y}_t | \mathbf{y}_{1:(t-1)})} \\
\text{Smoothing} \quad p(\mathbf{x}_t | \mathbf{y}_{1:n}) &= \int \frac{p(\mathbf{x}_t | \mathbf{y}_{1:t}) p(\mathbf{x}_{t+1} | \mathbf{x}_t)}{p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t})} p(\mathbf{x}_{t+1} | \mathbf{y}_{1:n}) d\mathbf{x}_{t+1}.
\end{aligned}$$

For ARMA(p, q) models: let $m = \max\{p, q + 1\}$, and

$$\begin{aligned}
Y_t &= \mathbf{F}_t \mathbf{X}_t + V_t \\
\mathbf{X}_{t+1} &= \mathbf{G}_t \mathbf{X}_t + \mathbf{W}_t
\end{aligned}$$

with state $\mathbf{X}_t = (X_{1,t-m+1}, X_{1,t-m+2}, \dots, X_{1,t})^\top$, and $\mathbf{W}_t = (0, 0, 0, \dots, Z_{t+1})^\top$, so that

$$\mathbf{G}_t = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ \phi_m & \phi_{m-1} & \phi_{m-2} & \cdots & \phi_2 & \phi_1 \end{bmatrix} \quad \mathbf{F}_t = \begin{bmatrix} \theta_{m-1} & \theta_{m-2} & \cdots & \theta_1 & \theta_0 \end{bmatrix}.$$

with $V_t \equiv 0$ and $Z_t \sim \text{Normal}(0, \sigma_Z^2)$.

The components of K -vector $\mathbf{X}_t$ are cointegrated of order (d, b) , $\mathbf{X}_t \sim CI(d, b)$ if the components of $\mathbf{X}_t$ are $I(d)$ (stationary after d -times differencing), and there exists a linear combination of $\mathbf{X}_t, \tilde{X}_t$, such that $\tilde{X}_t \sim I(d - b)$ for some $b > 0$.

If $\{Y_t\}, \{\mathbf{X}_t\} = \{(X_{1,t}, \dots, X_{K,t})^\top\} \sim I(1)$ are cointegrated, we fit the regression model

$$Y_t = \alpha_1 X_{1,t} + \dots + \alpha_K X_{K,t} + Z_t^*$$

with cointegrating vector $\alpha = (1, -\alpha_1, \dots, -\alpha_K)^\top$ and take residuals $\hat{z}_t^*$ should be stationary, which can be tested using the Dickey-Fuller procedure.

If

$$X_t^* = (1 - B)X_t \quad Y_t^* = (1 - B)Y_t$$

we may fit the error-correction models

$$\begin{aligned}
Y_t^* &= \psi_0 + \gamma_1 \hat{z}_{t-1}^* + \sum_{j=1}^p \psi_{1j} X_{t-j}^* + \sum_{l=1}^m \psi_{2l} Y_{t-l}^* + \epsilon_{1,t} \\
X_t^* &= \xi_0 + \gamma_2 \hat{z}_{t-1}^* + \sum_{j=1}^p \xi_{1j} Y_{t-j}^* + \sum_{l=1}^m \xi_{2l} X_{t-l}^* + \epsilon_{2,t}
\end{aligned}$$

to exploit the long-run relationship between $\{X_t\}$ and $\{Y_t\}$ for predicting changes in these series.

- **Spectral representations:** The spectral density function (SDF) of a stationary process $\{X_t\}$ is the function $f_X(\cdot)$ defined by

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\lambda h} \gamma_X(h) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \cos(\lambda h) \gamma_X(h) \quad \lambda \in (-\pi, \pi].$$

If F_X is the spectral distribution function, the inversion formula is

$$\gamma_X(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f_X(\lambda) d\lambda = \int_{-\pi}^{\pi} e^{ih\lambda} dF_X(\lambda).$$

For a linear process with coefficients given by $\Psi(z)$, we have

$$f_X(\lambda) = \frac{\sigma_Z^2}{2\pi} |\Psi(e^{-i\lambda})|^2 = \frac{\sigma_Z^2}{2\pi} \Psi(e^{-i\lambda}) \Psi(e^{i\lambda}).$$

An estimate of $2\pi f_X(\lambda)$ is the periodogram

$$\hat{I}_n(\lambda) = \frac{1}{n} \left| \sum_{t=1}^n x_t e^{-it\lambda} \right|^2 \quad -\pi < \lambda \leq \pi.$$

where for n large, and for $0 < \lambda_1 < \dots < \lambda_m < \pi$,

$$(\hat{I}_n(\lambda_1), \dots, \hat{I}_n(\lambda_m))^{\top} \xrightarrow{d} (V_1, \dots, V_n)^{\top}$$

as $n \rightarrow \infty$, $V_j \sim \text{Exponential}(1/(2\pi f_X(\lambda_j)))$

Long memory models are often constructed using fractionally integrated white noise: if $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and $W_t = (1 - B)^{-d} Z_t$ for $-0.5 < d < 0.5$, then

$$f_W(\lambda) = \frac{\sigma_Z^2}{2\pi} \frac{1}{|1 - e^{-i\lambda}|^{2d}}$$

- **Financial time series models:** Most models represent the volatility of the zero-mean observed process $\{Y_t\}$ by assuming that $Y_t = \sigma_t Z_t$ where $\{Z_t\}$ is a Gaussian white noise series, and

$$\begin{aligned} \text{ARCH}(p) \quad \sigma_t^2 &= \alpha_0 + \sum_{j=1}^p \alpha_j y_{t-j}^2 \\ \text{GARCH}(p, q) \quad \sigma_t^2 &= \alpha_0 + \sum_{j=1}^p \alpha_j y_{t-j}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \\ \text{SV} \quad \sigma_t &= \exp\{X_t/2\} \quad \text{with} \quad \{X_t\} \sim \text{AR}(1) \\ X_t &= \phi X_{t-1} + W_t \quad \{W_t\} \sim \text{WN}(0, \sigma_W^2). \end{aligned}$$

Stationarity properties of $\{Y_t\}$ depend on the values of the parameters: specifically, for the ARCH/GARCH model, we require

$$0 < \sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j < 1$$

and for the SV model we require that $\{X_t\}$ is stationary.