# Time Series Cheat Sheet

## Chapter 2: Basics

1. **Stationarity:** A discrete time stochastic process is called strongly stationary if the joint cdf of $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$ is the same as the joint cdf of $\{X_{t_1+s}, X_{t_2+s}, \ldots, X_{t_n+s}\}$. Weak stationarity means that the joint moments of order 1 and 2 of $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$ exist and are finite and are equal to the the joint moments for $\{X_{t_1+s}, X_{t_2+s}, \ldots, X_{t_n+s}\}$ for any valid $s$. Equivalently, we can say a process is second order stationary if the mean is constant and if the covariance is only dependent on lag $\tau$.

2. **White Noise:** A process which has constant mean, variance and no autocovariance.

3. **GLP:** A general linear process is one that depends only on the present and past and is given by

$$X_t = \sum_{k=0}^{\infty} g_k \epsilon_{t-k} = \left( \sum_{k=0}^{\infty} g_k B^k \right) \epsilon_t$$

where $\epsilon_t$ is a white noise process and it can be easily shown that this process is stationary and the autocovariance sequence of a GLP is given by $s_\tau = \sigma_\epsilon^2 \sum_{k=0}^{\infty} g_k g_{k+\tau}$.

4. **MA(q):** A moving average process given by

$$X_t = \mu - \theta_{0,q}\epsilon_t - \theta_{1,q}\epsilon_{t-1} - \cdots - \theta_{q,q}\epsilon_{t-q} = \mu - \sum_{i=0}^{q} \theta_{i,q}\epsilon_{t-i} = \mu + \Theta(B)\epsilon_t$$

and

$$\Theta(B) = 1 - \theta_{1,q}\epsilon_{t-1}B^1 - \cdots - \theta_{q,q}\epsilon_{t-q}B^q = 1 - \sum_{i=1}^{p} \theta_{i,p}B^i$$

is known as the characteristic polynomial. A moving average process is always stationary because it can be trivially written as a GLP. Furthermore, it is invertible if roots of $\Theta(z)$ outside $|z| \leq 1$.

5. **AR(p):** The autoregressive process is given by

$$X_t = \epsilon_t + \phi_{1,p}X_{t-1} + \cdots + \phi_{p,p}X_{t-p} \quad \text{or} \quad \Phi(B)X_t = \epsilon_t$$

where

$$\Phi(B) = 1 - \phi_{1,p}B - \cdots - \phi_{p,p}B^p = 1 - \sum_{i=1}^{p} \phi_{i,p}B^i$$

is known as the associated polynomial. This process is automatically invertible and also it is stationary if the roots of $\Phi(z)$ lie outside $|z| \leq 1$.

6. **ARMA(p,q)** is given by

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

and this is stationary or invertible if the roots of $\Phi(z)$ or $\Theta(z)$ lie outside $|z| \leq 1$.

7. **Trends and Seasonality:**

8. **Common Tricks**:

   a. Remembering that $\text{Cov}(\epsilon_t, \epsilon_{t'}) = 0$ for $t \neq t'$ to collapse the sum for calculation of autocovariance then setting changing the indexing.

   b. Using the geometric series forms for inverting polynomials.

## Chapter 3: Spectral Analysis

1. **Spectral Representation Theorem:** For real valued discrete time **stationary** process, there exists an orthogonal process $\{Z(f)\}$ on $[-1/2, 1/2]$ such that $X_t = \int_{-1/2}^{1/2} \exp(2\pi f t i)\, dZ(f)$. This is like a fourier transform where $Z(f)$ is the "fourier" transform of $X_t$. The process has the following properties

   a. $E\{dZ(f)\} = 0$ for all $|f| \leq 1/2$

   b. $E\{|dZ(f)|^2\} = dS^{(I)}(f)$ and $S^{(I)}$ is the integrated spectrum of $X_t$

   c. For $f \neq f' \in (-1/2, 1/2]$, $\text{Cov}(dZ(f), dZ(f')) = 0$

2. **Autocovariance:** We can relate the autocovariance sequence with the integrated spectrum

$$s_\tau = \int_{-1/2}^{1/2} \exp(2\pi fti)\, dS^{(I)}(f) = \int_{-1/2}^{1/2} \exp(2\pi fti) S(f)\, df$$

where the last equality holds only if the integrated spectrum is differentiable everywhere. Furthermore, we see that

$$S(f) = \sum_{\tau=-\infty}^{\infty} s_\tau \exp(-2\pi fti)$$

and hence we see that $S(f)$ is the fourier transform of $s_\tau$ and $s_\tau$ is the inverse fourier transform of $S(f)$ and they form a fourier pair. An easy way to remember is that the fourier transform is in the frequency domain.

3. **SDF**: Assuming the SDF exists it has the following properties

   a. FTC type $S^{(I)}(f) = \int_{-1/2}^{f} S(f')\, df'$

   b. Integrated spectrum is monotonous

   c. SDF is symmetric

4. **Classification of Spectra:** The integrated spectrum can be broken down into absolutely continuous and discrete parts where

$$S^{(I)}(f) = S^{(I)}_{ac}(f) + S^{(I)}_{d}(f) \quad \text{where} \quad S^{(I)}_{ac}(f) = \int_{-1/2}^{f} S(f')\, df \text{ and } S^{(I)}_{d}(f) \text{ is a step function}$$

5. **Sampling and Aliasing:** Given a sample rate, we need to determine what range of frequencies to integrate over to get $X(t)$. This leads to the defintion of the Nyquist frequency, $f_{\mathcal{N}} = 1/2\Delta t$, and we have the spectral representation theorems

$$X_t = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} e^{2\pi i f t \Delta t}\, dZ(f) \text{ and } X(t) = \int_{-\infty}^{\infty} e^{2\pi i f t}\, dZ(f)$$

where often discrete time series are continuous time series sampled at a constant frequency $\Delta t$. Because of this sampling we observe aliasing where The connection between discrete time and continuous time processes is that

$$S_{X_t}(f) = \sum S_{X(t)}\left(f + \frac{k}{\Delta t}\right) \text{ for } |f| \le 1/2\Delta t$$

6. **Linear Filtering:** A linear filter is a function from one sequence to another sequence that has the following properties

   a. $L(a(x_n)) = aL((x_n))$

   b. $L((x_1) + (x_2)) = L(x_1) + L(x_2)$

   c. $L(B^\tau(x_n)) = B^\tau L((x_n))$

   Then we also have $y_{f,t} = e^{i2\pi ft} y_{f,0}$ and we set $G(f) = y_{f,0}$ where $|G(f)|$ is the gain and $\arg(G(f))$ is the phase. Importantly, any linear filter can be represented as

$$L(X_t) = \sum_{u=-\infty}^{\infty} g_u X_{t-u} \equiv Y_t \quad \text{and}$$

$$L(e^{i2\pi ft}) = \sum_{u=-\infty}^{\infty} g_u e^{i2\pi f(t-u)} = e^{i2\pi ft} G(f) \implies G(f) = \sum_{u=-\infty}^{\infty} g_u e^{-i2\pi fu} \implies G(f) \leftrightarrow \{g_u\}$$

   We also have

$$dZ_Y(f) = G(f) dZ_X(f) \text{ and } E(|dZ_Y(f)|^2) = |G(f)|^2 E(|dZ_X(f)|^2) \text{ and } S_Y(f) = |G(f)|^2 S_X(f)$$

   if the spectral density function exists. The actual applications of this is by convolving the signal with the inverse fourier transform of a desired frequency response function we can filter out the undesired frequencies. To **find** $G(f)$, we can feed in a $e^{i2\pi ft}$ into the linear filter and the coefficient of $e^{i2\pi ft}$ would be the transfer/response function. We could also find the fourier transform of the impulse response sequence.

7. **Spectral Density using Linear Filters:** By **passing through** $L\{\xi_{f,t}\}$ to get $G(f)$, we can convert the spectral density of $X$ to the spectral density of $Y$ and then subsequently inverse fourier tranform the spectral density function to the autocovariance sequence.

   a. **MA:** For $X_t = \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q}$, we have $X_t = L(\epsilon_t) = \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q}$ and hence $G(f) = \Theta(e^{-i2\pi f})$ and $S_X(f) = |\Theta(e^{i2\pi ft})|^2 S_\epsilon(f) = |\Theta(e^{-i2\pi f})|^2 \sigma_\epsilon^2$.

   b. **AR:** For $X_t - \phi_{p,1}X_{t-1} - \ldots - \phi_{p,p}X_{t-p} = \epsilon_t$, we have $L(X_t) = X_t - \phi_{p,1}X_{t-1} - \ldots - \phi_{p,p}X_{t-p} = \epsilon_t$ and $G(f) = \phi(e^{-i2\pi f})$ and so $|\Phi(e^{-i2\pi f})|^2 S_X(f) = S_\epsilon(f) = \sigma_\epsilon^2$

8. **Main Tricks:**

a. Remember that $E(dZ^*(f)dZ(f)) = S(f)df$.

b. Remember that $S(f)$ is the fourier transform of $s_\tau$ so we can calculate the spectral density by first calculating the autocovariance sequence.

c. Remember the folding trick.

d. For ARMA, Linear filtering, remember that the trick is to get $\Phi(B)X_t = Y_t = \Theta(B)\epsilon_t$ then use lienar filtering to calculate the frequency response function.

## Chapter 4: Estimation

1. **MSE** $= E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + (bias(\hat{\theta}))^2$ and if MSE goes to 0, we have that the variance and bias goes to 0.

2. **Sample Mean**: converges in MSE to true mean when the autocovariance sequence is absolutely summable.

3. **Autocovariance:** A natural estimator for autocovariance is

$$\hat{s}_\tau^{(u)} = \frac{1}{N - |\tau|} \sum_{t=1}^{N-|\tau|} (x_t - \bar{x})(x_{t+|\tau|} - \bar{x})$$

which is unbiased when when we do not have to estimate $\bar{x} = \mu$. However, this might not be the best estimator and a second candidate is

$$\hat{s}_\tau^{(p)} = \frac{1}{N} \sum_{t=1}^{N-|\tau|} (x_t - \bar{x})(x_{t+|\tau|} - \bar{x})$$

and this is often preferred because the **MSE is lower**, the **autocovariance goes to 0** as $\tau \to N - 1$ and because $\hat{s}_t^{(p)}$ is a **positive semidefinite** sequence whereas $\hat{s}_t^{(u)}$ might not be.

4. **Spectral Density Estimation:** For spectral density estimation we just do the fourier transform of the autocovariance sequence estimate to get

$$\hat{S}^{(p)}(f) = \sum_{\tau=-(N-1)}^{N-1} \hat{s}_\tau^{(p)} e^{-2i\pi f t} = \frac{1}{N} \left| \sum_{t=1}^{N} X_t e^{-2i\pi f t} \right|^2$$

where this is just the normalised square of the fourier transform of the time series. This is known as the periodogram/ power spectrum. However, this is a good approximation for some processes and bad for others, its variance does not go to 0 as $N \to \infty$ and lastly, $\mathrm{Cov}(\hat{S}^{(p)}(f), \hat{S}^{(p)}(f')) \approx 0$ if $f, f'$ are fourier frequencies.

5. **Expected Value of Periodogram**: By using the fact that the estimator $\hat{S}^{(p)}(f)$ is the normalised square of the fourier transform of the time series, we see that the expected value of the periodogram is given by

$$E(\hat{S}^{(p)}(f)) = \int_{-1/2}^{1/2} \mathcal{F}(f - f')S(f')df' = \mathcal{F} * S \text{ where } \mathcal{F} = \frac{\sin^2(N\pi f)}{N \sin^2(\pi f)}$$

Because the convolution of another function with a delta function recovers the original function, we want the Fejers kernel to be close to a dirac delta function. So the Fejers kernel approaches $N$ as $f \to 0$, for any frequency not 0, $\mathcal{F}(f) \to 0$ pointwise, it achieves its maximum at 0, for any fourier frequency, $f_k$, $\mathcal{F}(f_k) = 0$ and integrates to 1 on $[-1/2, 1/2]$. Therefore, as $N \to \infty$, $\mathcal{F} \to \delta$ and $E(\hat{S}^{(p)}(f)) \to S(f)$, ie asymptotic unbiased.

6. **Bias tapering:** The effects of side lobe leakage is worse for data with large frequency domain. A data taper is a sequence of the length N such that $\sum_i h_i^2 = 1$. We can the fourier transform $h_t X_t$ and get

$$\hat{S}^{(d)}(f) = \left| \sum_{t=1}^{N} h_t X_t e^{-2i\pi f t} \right|^2 \text{ and } E(\hat{S}^{(d)}(f)) = \int_{-1/2}^{1/2} \mathcal{H}(f - f')S(f')df' \text{ where } \mathcal{H}(f) = \left| \sum_{t=1}^{N} h_t e^{-i2\pi f t} \right|^2$$

Ideally we want to choose $h_t$ such that $\mathcal{H}$ is close to a delta function and also if $h_t = 1/\sqrt{n}$ then $\hat{S}^{(d)}(f) = \hat{S}^{(p)}(f)$. The downfall is that the resolution of the spectral estimate is worse (ie. wider central lobe).

7. **Common Tricks:**

   a. Remember the sum order swapping from diagonal sums to row sums.

   b. Comparing coefficients to recover the original characteristic polynomial.

## Chapter 5: Model Fitting (AR Models)

1. **Yule Walker:** The Yule Walker method leverages that fact that

$$s_\tau = \sum_{j=1}^{p} \phi_{j,p}\, s_{\tau-j} \quad \text{and} \quad \hat{\sigma}_\epsilon^2 = \hat{s}_0 - \sum_{j=1}^{p} \hat{\phi}_{j,p} \hat{s}_j$$

so that we have $\hat{\phi}_p = \hat{\Gamma}^{-1}\hat{\gamma}_p$ where $\hat{\Gamma}^{-1}$ is the estimated Toeplitz matrix and $\hat{\gamma}_p$ is the estimated autocovariance sequence. Furthermore, using the formulation for SDFs based on linear filtering we have

$$\hat{S}(f) = \frac{\hat{\sigma}_\epsilon^2}{\left|\hat{\Phi}(e^{i2\pi f t})\right|^2}$$

And Yule Walker always gives a stationary sequence if using the biased estimator $\hat{s}^{(p)}$.

2. **LS:** Just the usual least squares where the covariates are the past p time points and the dependent is the current time point. We can find by minimising sum of squared errors to get

$$\hat{\phi} = (F^T F)^{-1} F^T X \quad \text{and} \quad \hat{\sigma}_\epsilon^2 = \frac{(X - F\hat{\phi})^T (X - F\hat{\phi})}{N - 2p}$$

Additionally, if you append 0s behind to $1-p$ and 0s ahead to $N+p$ and the calculate the LS estimate, it is equivalent to the Yule-Walker estimate.

3. **MLE:** This is equivalent to LS approach when we assume the process is Gaussian. But can be maximised by using

$$f(X_1, X_2, \ldots, X_n | \phi, \sigma_\epsilon^2) = f(X_1, X_2, \ldots, X_n) \prod_{t=p+1}^{N} f(X_t | X_{t-1}, X_{t-2}, \ldots, X_{t-p}, \phi, \sigma_\epsilon^2)$$

4. **Model Selection:** You choose the model which has the least number of parameters but which is still effective and the criterion is given by

$$\text{AIC} = 2k - 2\ln(\ell(\hat{\phi}, \sigma_\epsilon^2))$$

where $k = p + 1$ and $\ell(\cdot,\cdot)$ is the liklihood function.

## Chapter 6: Forecasting

a. **Formulation:** For a stationary and invertible ARMA process (because we need GLP) we have $X_{t+l} = \Psi(B)\epsilon_{t+l}$ where $\Psi$ is the GLP. Next clip the part dependent on future because the expected value of their contributions are 0 to get

$$X_t(l) = \sum_{k=0}^{\infty} \psi_{k+l}\epsilon_{t-k} = \Psi^{(l)}(B)\epsilon_t$$

This forecast minimises the expected squared difference between the forecast and the actual representation. We also have the l-step Ahead Prediction Variance

$$\sigma^2(l) = E((X_{t+l} - X_t(l))^2) = \sigma_\epsilon^2 \sum_{k=0}^{l-1} \psi_k^2$$

b. **Inverting:** Assuming $X_t$ is also invertible we see

$$\epsilon_t = \Psi^{-1}(B)X_t \implies X_t(l) = \Psi^{(l)}(B)\Psi^{-1}(B)X_t = G^{(l)}(B)X_t$$

c. **Error:**

$$e_t(l) = X_{t+l} - X_t(l) = \sum_{k=0}^{l-1} \psi_k \epsilon_{t+l-k} \quad \text{and} \quad \text{Cov}(e_t(l), e_t(m+l)) = \sigma_\epsilon^2 \sum_{k=1}^{l-1} \psi_k \psi_{k+m}$$

d. **Update:**

$$X_{t+1}(l) = X_t(l+1) + \psi_l(X_{t+1} - X_t(1))$$

## Chapter 7: Bivariate

a. **Joint Stationarity:** Two real valued discrete stochastic processes are jointly stationary stochastic process if both $X_t, Y_t$ are each weakly stationary and cross covariance is dependent only on lag $\tau$.

b. **Cross Covariance:** The first is at time $t$ and the second is at time $t + \tau$

$$s_{X,Y,\tau} = E((X_t - \mu_X)(Y_{t+\tau} - \mu_Y))$$

Furthermore, not necessarily $s_{X,Y,\tau} = s_{X,Y,-\tau}$ and the sequence is generally assymetric.

c. **Estimating CCV:** We can estimate this similar to the biased estimator $\hat{s}_\tau^{(p)}$ from before

d. **Linear Filtering w Noise Example:** For 0 mean stationary $X_t, Y_t$ where $Y_t = \sum_{u=-k}^{u=k} g_u X_{t-u} + \eta_t$ we have a useful way to get the cross covariance sequence

$$s_{X,Y,\tau} = E\left(X_t\left(\sum_{u=-k}^{u=k} g_u X_{t-u} + \eta_t\right)\right) = \sum_{u=-k}^{u=k} g_u s_{X,\tau-u} = g * s_X$$

e. **Cross Spectra:** Just the fourier transform of the cross covariance sequence and also no longer real value because ccv sequence is not symmetric so complex parts do not cancel.

$$S_{X,Y}(f) = \sum_{\tau=-\infty}^{\infty} s_{X,Y,\tau}\, e^{-i2\pi f\tau} \text{ for } |f| \leq 1/2$$

We do have $S_{X,Y}^*(f) = S_{X,Y}(-f)$ ie skew symmetric. Furthermore, if the two processes are **cross orthogonal**, we also have that

$$E(dZ_X(f)dZ_Y(f')) = \begin{cases} S_{X,Y}(f)\, df & \text{if } f = f' \\ 0 & \text{otherwise} \end{cases}$$

and hence we see

$$s_{X,Y,\tau} = E(X_t^* Y_{t+\tau}) = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{-i2\pi ft} e^{i2\pi f'(t+\tau)} E(dZ_X(f)dZ_Y(f')) = \int_{-1/2}^{1/2} e^{-i2\pi f\tau} S_{X,Y}(f)\, df$$

and we have the **spectral matrix** which is Hermitian

$$S(f) = \begin{pmatrix} S_X(f) & S_{X,Y}(f) \\ S_{Y,X}(f) & S_Y(f) \end{pmatrix}$$

f. **Coherence:** Kinda like $R^2$ and is given by

$$\gamma_{X,Y}^2(f) = \frac{|S_{X,Y}(f)|^2}{S_X(f)S_Y(f)}$$

g. **Linear Filter + Noise Model:** Like above we also have

$$S_{X,Y}(f) = G(f)S_X(f) \text{ and } S_Y(f) = |G(f)|^2 S_X(f) + S_\eta(f)$$

h. **Bivariate AR**: We get

$$\Phi(B)X_t = \epsilon_t \text{ where } \Phi(\mathbf{z}) = 1 - \phi_{1,p}z - \cdots - \phi_{p,p}x^p \text{ where } \phi_{i,p} \in \mathbb{R}^{2\times 2}$$

and $\epsilon_t$ is a bivariate white noise process and $E(\epsilon_t^T e_s) = \Sigma$ if $s = t$ and this is stationary if roots of determinant of $\phi(\mathbf{z})$ are outside 1.