Time Series lecture 14

Revisions

Sofia Olhede

EPFL

May 28, 2025

Lecture outline

1. Definitions
2. Models
3. Frequency domain time series
4. Linear Time Invariant filters
5. Multivariate time series
6. Estimation
7. Forecasting

Definitions

- Informally, a time series X_t is just data recorded over time.
- We shall use the word 'time series' to mean both the data, and the process from which the data is a realisation.
- More formally, we think of X_t as a stochastic process, i.e. as a family of random variables $\{X_t : \ t \in \mathcal{T}\}$ defined on a probability space $(\Omega, \mathscr{F}, P)$.
- In time series analysis the index (or parameter) set $\mathcal{T}$ is a set of time points, very often $\mathbb{R}$ or $\Delta\mathbb{Z}$ (or a subset of them).
- Here $\Delta \in \mathbb{R}$ is the time step between observations.

Definition 14.1 ((Weak) Stationarity)

The time series $\{X_t\}$ is said to be second-order/weak or covariance stationary if for all $n \geq 1$ for any $t_1, \ldots, t_n \in \mathcal{T}$ and for all τ such that $t_1 + \tau, \ldots, t_n + \tau \in \mathcal{T}$ all the joint moments of order 1 and 2 of $X_{t_1}, \ldots, X_{t_n}$ exist, are all finite and equal to the corresponding joint moments of $X_{t_1+\tau}, \ldots, X_{t_n+\tau}$.

In fact this corresponds to that, for all $t, s, \tau \in \mathcal{T}$

1. $\mathbb{E}[X_t] = \mu$,
2. $\text{Var}(X_t) = \sigma^2 < \infty$,
3. $\mathbb{E}[X_t X_{t+\tau}] = \mathbb{E}[X_s X_{s+\tau}]$.

One may deduce from this that $\mathbb{E}[X_t X_{t+\tau}]$ can be written as a function of τ only.

We can go beyond the first two moments and define strong stationarity.

Definition 14.2 (Strong Stationarity)

The time series $\{X_t\}$ is said to be completely/strong or strictly stationary if for all $n \geq 1$ for any $t_1, \ldots, t_n \in \mathcal{T}$ and for all τ such that $t_1 + \tau, \ldots, t_n + \tau \in \mathcal{T}$ the joint distribution of $X_{t_1}, \ldots, X_{t_n}$ is the same as $X_{t_1+\tau}, \ldots, X_{t_n+\tau}$.

In general:

- second order stationary $\not\Rightarrow$ strictly stationary,
- strict stationarity $\not\Rightarrow$ 2nd order stationarity.

Definition 14.3 (ACVS)

For a discrete time second-order stationary process $\{X_t\}$ we define the autocovariance sequence (ACVS) by

$$\gamma_\tau = \mathrm{Cov}\left(X_0, X_\tau\right), \tag{14.1}$$

where $\tau \in \mathbb{Z}$ is the lag.

Definition 14.4 (ACF)

The autocorrelation sequence, usually called autocorrelation function, (ACF) is defined as

$$\rho_\tau = \mathrm{Corr}\left(X_t, X_{t+\tau}\right), \tag{14.2}$$

for $\tau \in \mathbb{Z}$.

- We have $\rho_\tau = \gamma_\tau / \gamma_0$.

Models

Definition 14.5 (MA(q))

Let $\{\varepsilon_t\}$ be a mean-zero white noise process. Then we define the <u>q-th order moving average process</u>, denoted MA(q), as

$$X_t = \mu - \sum_{j=0}^{q} \theta_j \varepsilon_{t-j}, \tag{14.3}$$

where θ_j are constants such that $\theta_0 = -1$ and $\theta_q \neq 0$.

- The only constraint for stationarity is that $|\theta_j| < \infty$.

Definition 14.6 (AR(p))

Let $\{\varepsilon_t\}$ be a mean-zero white noise process. Then we define the <u>p-th order autoregressive process</u>, denoted AR(p), as

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t, \tag{14.4}$$

where ϕ_j are constants such that $\phi_p \neq 0$.

- The requirement for stationarity is that the roots of $\Phi(z)$ are outside the unit disc.

Definition 14.7 (ARMA(p, q))

A time series $\{X_t\}$ is an autoregressive moving average process of order p and q, denoted ARMA(p, q), if

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} - \sum_{k=0}^{q} \theta_k \varepsilon_{t-k}$$

where $\{\varepsilon_t\}$ is a mean-zero white noise process, and ϕ_j, θ_k are the same as in the AR and MA cases respectively.

Often an observed signal exhibits a trend. This is a tendency to increase or decrease over time. There may also be fluctuations over time. This model is given by

$$X_t = \mu_t + Y_t,$$

where μ_t is a time-dependent mean, and Y_t is a stationary process, for example $\mu_t = a + bt$.

- First differences $\Delta X_t = X_t - X_{t-1}$.
- In fact for the first difference of a stationary process is stationary, so if Y_t was stationary then so is ΔY_t.
- It is convenient to define the backshift operator B. $BX_t = X_{t-1}$.
- If we difference again then we arrive at

$$\begin{aligned}\Delta^2 X_t &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= X_t - 2X_{t-1} + X_{t-2} \\ &= \Delta Y_t - \Delta Y_{t-1} \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2}.\end{aligned}$$

The general linear process takes the form

$$X_t = \sum_{k=0}^{\infty} g_k \epsilon_{t-k}.$$

Theorem 14.8 (The Wold Decomposition Theorem)

Any stationary process X_t can be expressed in the form $X_t = U_t + V_t$ where

1. *U_t and V_t are uncorrelated processes*
2. *U_t has the one-sided representation $U_t = \sum_u g_k \epsilon_{t-u}$*
3. *V_t is singular*

- Consider inverting a process

$$X_t = G(B)\epsilon_t \Rightarrow \epsilon_t = G^{-1}(B)X_t.$$

  The expansion of $G^{-1}(B)$ in powers of B gives the AR form, provided $G^{-1}(B)$ admits a power expansion.
- Thus the model is invertible if $|G^{-1}(z)| < \infty$ for $|z| \leq 1$.
- An AR $\Phi(B)X_t = \epsilon_t$ is stationary if the roots of $\Phi(z)$ are outside the unit circle.

When are you stationary/invertible?

- What does this mean in practice? We have treated General Linear Processes!
- Assume we have an AR process (ϵ_t is assumed to be a white noise process):

$$X_t + \frac{1}{10}X_{t-1} - \frac{48}{100}X_{t-2} = \varepsilon_t.$$

- We can write this with the backshift operator

$$\Phi(B) = I + \frac{1}{10}B - \frac{48}{100}B^2.$$

- Solve $\Phi(z) = I + \frac{1}{10}z - \frac{48}{100}z^2 = 0$.
- This has solutions

$$z_0 = \frac{1}{10}\frac{50}{48} \pm \frac{\sqrt{\frac{1}{100} + 4 \times \frac{48}{100}}}{48/50} = \frac{5}{48} \pm \frac{5\sqrt{193}}{48} = 1.55,\ -1.34.$$

- Stationary and invertible.

When are you stationary/invertible?

- Assume ϵ_t is a white noise process with variance σ^2. We then define the moving average process by

$$X_t = \varepsilon_t - 2\varepsilon_{t-1}.$$

- Is this a stationary process? Yep MAs are always stationary.
- Is this an invertible process? We define $\theta(z) = 1 - 2z$. No, the root is $1/2$. This is not outside the unit circle.
- Can we define a process with the same auto-correlation? Note

$$\begin{aligned} \text{Var}\left(X_t\right) &= \text{Var}\left(\varepsilon_t - 2\varepsilon_{t-1}\right) = 5\sigma^2 \\ \text{Cov}\left(X_t, X_{t+\tau}\right) &= \text{Cov}\left(\varepsilon_t - 2\varepsilon_{t-1}, \varepsilon_{t+\tau} - 2\varepsilon_{t+\tau-1}\right) \\ &= \sigma^2\left\{\delta_{0,\tau} - 2\delta_{1,\tau} - 2\delta_{-1,\tau} + 4\delta_{0,\tau}\right\}. \qquad (14.5) \end{aligned}$$

When are you stationary/invertible?

- Assume ϵ_t is a white noise process with variance σ^2. We then define the moving average process by

$$X_t = \varepsilon_t - \frac{1}{2}\varepsilon_{t-1}.$$

- Is this a stationary process? Yep MAs are always stationary.
- Is this an invertible process? We define $\theta(z) = 1 - \frac{1}{2}z$. Yes, the root is 2. This is outside the unit circle.
- Note

$$\begin{aligned} \mathrm{Var}\left(X_t\right) &= \mathrm{Var}\left(\varepsilon_t - 2^{-1}\varepsilon_{t-1}\right) = 5\sigma^2/4 \\ \mathrm{Cov}\left(X_t, X_{t+\tau}\right) &= \mathrm{Cov}\left(\varepsilon_t - 2^{-1}\varepsilon_{t-1}, \varepsilon_{t+\tau} - 2^{-1}\varepsilon_{t+\tau-1}\right) \\ &= \sigma^2\left\{\delta_{0,\tau} - 2^{-1}\delta_{1,\tau} - 2^{-1}\delta_{-1,\tau} + 2^{-2}\delta_{0,\tau}\right\}. \quad (14.6) \end{aligned}$$

When are you stationary/invertible?

- Assume ϵ_t is a white noise process with variance σ^2. We then define the ARMA process by

$$X_t + \frac{3}{5}X_{t-2} = \varepsilon_t + \frac{6}{5}\varepsilon_{t-1}.$$

We determine the polynomials $\Phi(z) = 1 + \frac{3}{5}z^2$ and $\Theta(z) = 1 + \frac{6}{5}z$. This has roots $z^2 = -\frac{5}{3}$ and $z = 5/6$. Stationary then but not invertible.

Frequency domain time series

Let $\{X_t\}$ be a real-valued discrete time stationary process with mean μ. As a result of Herglotz's theorem,

$$\gamma_\tau = \int_{-1/2}^{1/2} e^{2\pi i \tau f} \mathrm{d}S^{(I)}(f)\,. \tag{14.7}$$

Also, there exists an orthogonal increment process $\{Z(f)\}$ on $\left[-\frac{1}{2}, \frac{1}{2}\right]$ such that

$$X_t = \mu + \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i f t} \mathrm{d}Z(f). \tag{14.8}$$

This equality holds in the mean–square sense. The process $\{Z(f)\}$ has properties for $f, f' \in \left[-\frac{1}{2}, \frac{1}{2}\right]$

1. $\mathbb{E}\left[\mathrm{d}Z(f)\right] = 0$,
2. $\mathrm{Var}\left(\mathrm{d}Z(f)\right) = \mathrm{d}S^{(I)}(f)$,
3. $\mathrm{Cov}\left(\mathrm{d}Z(f), \mathrm{d}Z(f')\right) = 0$ if $f \neq f'$.

- If $\gamma \in \ell^1$, the spectral density function is

$$S(f) = \sum_{\tau \in \mathbb{Z}} \gamma_\tau e^{-2\pi i \tau f} \qquad (14.9)$$

for all $f \in \mathbb{R}$.

- In this case, this relates to the integrated spectrum by

$$S^{(I)}(f) = \int_{-1/2\pi}^{f} S(\lambda)\,\mathrm{d}\lambda. \qquad (14.10)$$

- The spectral density function also satisfies, for all $f \in \mathbb{R}$
  1. $S(f) \geq 0$,
  2. $S(f) = S(-f)$,
  3. $\int_{-1/2\pi}^{1/2\pi} S(\lambda)\,\mathrm{d}\lambda = \sigma^2$.

Consider a continuous time process $\{X_t\}_{t\in\mathbb{R}}$, with autocovariance function $\gamma \in L^1$, then the spectral density function is

$$\mathscr{S}(f) = \int_{-\infty}^{\infty} \gamma(\tau)\, e^{-2\pi i \tau f} \mathrm{d}\tau \tag{14.11}$$

for $f \in \mathbb{R}$.

- If we construct the discrete-time process $\{X_t\}_{t\in\mathbb{Z}}$ by sampling the continuous process, we get the aliasing relation

$$S(f) = \sum_{k\in\mathbb{Z}} \mathscr{S}(f+k)\,. \tag{14.12}$$

- The frequency $\frac{1}{2}$ is called the Nyquist frequency, after which $S(f)$ repeats.

Linear Time Invariant filters

- A digital filter L that transforms an input sequence into an output sequence is called a linear time invariant (LTI) digital filter if it satisfies scale preservation, superposition and time invariance.
- In this context we defined the transfer function, the frequency response function and the impulse response function.

Multivariate time series

$\{\mathbf{X}_t\}$ denotes a real d-vector-valued discrete time stochastic process with

$$\mathbf{X}_t = \begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \\ \vdots \\ X_t^{(d)} \end{pmatrix}$$

Second-order stationarity requires for all t

$$\mu_k = \mathbb{E}\left[X_t^{(k)}\right],$$
$$\gamma_\tau^{(k)} = \mathrm{Cov}\left(X_{t+\tau}^{(k)}, X_t^{(k)}\right) < \infty,$$
$$\gamma_\tau^{(k,l)} = \mathrm{Cov}\left(X_{t+\tau}^{(k)}, X_t^{(l)}\right)$$

to be independent of t.

- Bivariate processes were considered in more detail, including special forms of dependence such as processes with contemporaneous correlation, processes with contemporaneous correlation.
- These processes also admit a spectral representation. Their cross-covariance admits a cross-spectrum.

Estimation

- We can estimate the mean by $\widehat{\mu} = \overline{X}$. Additionally we have

$$\tilde{\gamma}_\tau = \frac{1}{N-|\tau|}\sum_{t=1}^{N-|\tau|}\{X_t - \overline{X}\}\{X_{t+|\tau|} - \overline{X}\}, \quad \tau = 0, \pm 1, \pm 2, \ldots.$$

$$\hat{\gamma}_\tau = \frac{1}{N}\sum_{t=1}^{N-|\tau|}\{X_t - \overline{X}\}\{X_{t+|\tau|} - \overline{X}\}, \quad \tau = 0, \pm 1, \pm 2, \ldots.$$

- To analyse this estimator we normally replace $\overline{X}$ with the population version.

- We then arrive at the spectral estimator via (when assuming $\mu = 0$)

$$\hat{S}^{(p)}(f) = \Delta t \sum_{\tau=-(N-1)}^{(N-1)} \hat{\gamma}_\tau(\tau) e^{-2i\pi f \tau \Delta t}$$
$$= \frac{\Delta t}{N} \left| \sum_{j=1}^{N} X_j e^{-2i\pi f j \Delta t} \right|^2 . \qquad (14.13)$$

This estimator is asymptotically unbiased, inconsistent, and approximately uncorrelated between special frequencies.

- We can define a tapered version of the periogram by taking

$$\hat{S}^{(p)}_{h_k}(f) = \left| \sum_{t=1}^{N} h_{t,k} X_t e^{-2i\pi ft} \right|^2 .$$

$\{h_{t,k}\}$ is a taper function.

- The multitaper estimator is for $K \geq 2$

$$\hat{S}^{(mt)}(f) = \frac{1}{K} \sum_{k=1}^{K} \hat{S}^{(p)}_{h_k}(f).$$

- We can revisit the autoregressive processes.
- We can obtain the Yule–Walker equations.
- Estimation is based on estimating the ACVS with and without tapering.
- We can estimate the AR parameters using the forward least squares estimator, the backwards least squares estimator, and the forward/backward least squares estimator.

- We also covered the Box Jenkins framework for modelling time series. This included identification, based on autocovariance and partial autocovariance plots.
- The framework included estimation, using least squares, rather than the Yule-Walker method of moments.
- The framework included model checking using residual plots, and Box Pierce statistics.
- Overfitting and model choice using information criteria such as Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC).

Forecasting

For $h > 0$, the minimum prediction mean square error forecast is attained by

$$X_{n+h}^n = \mathbb{E}\left[X_{n+h} \mid X_1, \ldots, X_n\right]$$

For AR processes forecasting is very natural. How do we then forecast a general ARMA?
We use the truncated predictions (see lecture 10). The residuals $\widehat{\varepsilon}_t$ and fitted values/predictions X_t^n are obtained recursively by

$$\widehat{\varepsilon}_t = \begin{cases} \sum_{i=1}^{p} \phi_i X_{t-i}^n + \left(\theta_1 \widehat{\varepsilon}_{t-1} + \cdots + \theta_q \widehat{\varepsilon}_{t-q}\right), & t = 1, \ldots, n, \\ 0, & \text{otherwise} . \end{cases}$$

$$X_{t+h}^n = \begin{cases} \sum_{i=1}^{p} \phi_i X_{t+h-i}^n - \sum_{j=1}^{q} \theta_j \widehat{\varepsilon}_{t+h-j}, & t+h > n, \\ X_{t+h}, & t+h = 1, \ldots, n, \\ 0, & t+h \leq 0. \end{cases}$$

- We can find the properties of the *step*-step ahead forecast error using an infinite MA representation. (Theorem 10.6)
- Various measures of prediction performance were introduced.