STAT 153 & 248 - Time Series
Lecture Nineteen

Fall 2025, UC Berkeley

Aditya Guntuboyina

November 04, 2025

We have discussed AR(p) models in the past few lectures. We discussed parameter estimation in AR(p) and also prediction or forecasting of future observations. We will make some more observations about AR models today. After that we will discuss Moving Average (MA) models, and more generally ARIMA models (which is a more general class of models including AR and MA models).

Today we first take a general look at time series models (especially the models that we have already considered), and introduce the concept of "Stationarity". Stationarity is an important property that some time series models satisfy while others do not.

1 Time Series Models and Stationarity

We have already seen many models for observed time series data $y_1, \dots, y_n$. These models describe the distribution of $y_1, \dots, y_n$ in terms of various parameters. Even though, the observed data only corresponds to times $t = 1, \dots, n$, it makes sense for the model to describe the distribution of y_t for all t past and present i.e., $t = \dots, -3, -2, -1, 0, 1, 2, 3, \dots$. This is because one may be interested in predicting the values of y_t for unobserved times.

All the models we will study have error terms ϵ_t that we assume are i.i.d Gaussian $N(0, \sigma^2)$. For all models that we have so far studied, the i.i.d Gaussian errors assumption ensures that the whole time series $\{y_t\}$ is jointly Gaussian (this will not be true for the neural network models that we will consider later). Gaussianity ensures that the behavior of the time series model is characterized by means (expectations) and covariances.

Some time series models satisfy the property of **stationarity** which is defined as follows:

Definition 1.1 (Stationarity). *A time series model for $y_t, t = \dots, -3, -2, -1, 0, 1, 2, 3, \dots$ is said to be stationary if both the following conditions hold:*

1. *The mean of y_t (denoted by $\mathbb{E}y_t$) is the same for all times t*
2. *The variance of y_t (denoted by $var(y_t)$) is the same for all times t*
3. *The covariance between y_{t_1} and y_{t_2} only depends on the distance $|t_1 - t_2|$ between t_1 and t_2.*

Stationarity implies, for example, that the mean of y_{-2000} should be the same as y_{9999}. Also the covariance between y_{-2000} and y_{-2100} should be the same as the covariance between y_{9899} and y_{9999} etc.

For a stationary time series model $\{y_t\}$, the covariance between y_t and y_{t+h} will only depend on $|h|$. We denote:

$$\gamma(h) = \text{cov}(y_t, y_{t+h}) \qquad \text{for } h = \ldots, -2, -1, 0, 1, 2, \ldots.$$

$\gamma(h)$ is called the AutoCovariance Function (ACVF) of the stationary time series model $\{y_t\}$. Observe that

$$\gamma(0) = \text{cov}(y_t, y_t) = \text{var}(y_t) \;\&\; \gamma(-h) = \text{cov}(y_t, y_{t-h}) = \text{cov}(y_{t-h}, y_t) = \text{cov}(y_{t-h}, y_{t-h+h}) = \gamma(h)$$

So $\gamma(h)$ is a symmetric function of h, and we only need to evaluate it at nonnegative h.

The AutoCorrelation Function (ACF) of a stationary time series model $\{y_t\}$ is defined as:

$$\rho(h) = \text{correlation between } y_t \text{ and } y_{t+h} = \frac{\text{cov}(y_t, y_{t+h})}{\sqrt{\text{var}(y_t)\text{var}(y_{t+h})}} = \frac{\gamma(h)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \frac{\gamma(h)}{\gamma(0)}.$$

Note that $\rho(0) = 1$ and $\rho(h) = \rho(-h)$.

It is important to remember the following points:

1. Stationarity refers to a time series model (not to actual data)
2. Not all time series models are stationary. In fact, there are many time series models that are not stationary.
3. ACVF and ACF are only defined for stationary time series models.

Let us now look at some examples of time series models, starting with the simplest.

Example 1.2 (Gaussian White Noise)**.** *The simplest time series model is* $y_t = \epsilon_t$ *where* $\epsilon_t \overset{i.i.d}{\sim} N(0, \sigma^2)$*. This is known as the Gaussian white noise model. It is easy to check that* $\mathbb{E}y_t = 0$ *and* $cov(y_t, y_{t+h}) = \sigma^2 I\{h = 0\}$*. So the conditions of stationarity are satisfied, and the Gaussian white noise is a stationary model. Its ACF is* $\rho(h) = I\{h = 0\}$*.*

Example 1.3 (Constant mean plus White Noise)**.** *The next simplest time series model is* $y_t = \mu + \epsilon_t$ *where, again,* $\epsilon_t \overset{i.i.d}{\sim} N(0, \sigma^2)$*. This is also a stationary time model because* $\mathbb{E}y_t = \mu$ *and* $cov(y_t, y_{t+h}) = \sigma^2 I\{h = 0\}$*. Its ACF is* $\rho(h) = I\{h = 0\}$*.*

The next two examples are for non-stationary time series models.

Example 1.4. $y_t = \beta_0 + \beta_1 t + \epsilon_t$*. Here the mean of* y_t *is:*

$$\mathbb{E}y_t = \beta_0 + \beta_1 t$$

and covariances are:

$$var(y_t) = \sigma^2 \qquad and \qquad cov(y_{t_1}, y_{t_2}) = 0 \;\; for\; t_1 \neq t_2.$$

So the mean changes with t*, variance is constant and there is no correlation between different time points. Because the mean changes with* t*, this is a non-stationary model.*

Example 1.5. $y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + \epsilon_t$ *The means are given by:*

$$\mathbb{E} y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$$

and covariances are:

$$var(y_t) = \sigma^2 \quad \text{and} \quad cov(y_{t_1}, y_{t_2}) = 0 \ \text{ for } t_1 \neq t_2.$$

Again the mean changes with t, variance is constant and there is no correlation between different time points. Because the mean changes with t, this is non-stationary.

The next example is the spectrum model.

Example 1.6. *Consider the Spectrum model:*

$$y_t = \beta_0 + \sum_{j=1}^{m} \left(\beta_{1j} \cos \frac{2\pi j t}{n} + \beta_{2j} \sin \frac{2\pi j t}{n} \right) \tag{1}$$

with $\beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}, \dots, \beta_{1m}, \beta_{2m}$ all independent with

$$\beta_{1j}, \beta_{2j} \overset{i.i.d}{\sim} N(0, \tau_j^2).$$

Here m is the largest positive integer that is strictly smaller than $n/2$.

For this model, $\mathbb{E} y_t = \beta_0$ so that the mean is constant over time (unlike the previous two models). The covariance is given by

$$\begin{aligned}
&cov(y_{t_1}, y_{t_2}) \\
&= cov\left(\beta_0 + \sum_{j=1}^{m} \left(\beta_{1j} \cos \frac{2\pi j t_1}{n} + \beta_{2j} \sin \frac{2\pi j t_1}{n} \right), \beta_0 + \sum_{j=1}^{m} \left(\beta_{1j} \cos \frac{2\pi j t_2}{n} + \beta_{2j} \sin \frac{2\pi j t_2}{n} \right) \right) \\
&= \sum_{j=1}^{m} \left\{ cov\left(\beta_{1j} \cos \frac{2\pi j t_1}{n}, \beta_{1j} \cos \frac{2\pi j t_2}{n} \right) + cov\left(\beta_{2j} \sin \frac{2\pi j t_1}{n}, \beta_{2j} \sin \frac{2\pi j t_2}{n} \right) \right\} \\
&= \sum_{j=1}^{m} \left\{ \tau_j^2 \cos \frac{2\pi j t_1}{n} \cos \frac{2\pi j t_2}{n} + \tau_j^2 \sin \frac{2\pi j t_1}{n} \sin \frac{2\pi j t_2}{n} \right\} \\
&= \sum_{j=1}^{m} \tau_j^2 \cos \frac{2\pi j (t_1 - t_2)}{n} = \sum_{j=1}^{m} \tau_j^2 \cos \frac{2\pi j |t_1 - t_2|}{n}
\end{aligned}$$

This model incorporates dependence between y_t at different time points t (unlike the previous two models). Further, the covariance between y_{t_1} and y_{t_2} only depends on the distance $|t_1 - t_2|$ between the two time points.

A slightly inelegant thing about this model is that the specification (1) depends on the sample size n. One can make the frequencies to take values in the continuum $(0, 1/2)$ and replace the sum $\sum_{j=1}^{m}$ by an integral. This can be done for example as in Shumway and Stoffer [1, Theorem C.2].

Our next example is AR models. Are AR models stationary? The answer is a bit complicated. Let us start with AR(1) and then consider AR(p) for $p \geq 2$ in the next lecture.

2 Stationarity of AR(1)

The AR(1) equation is

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t \tag{2}$$

One issue is that this equation does not fully specify y_t and there can be multiple processes $\{y_t\}$ that satisfy (2):

1. Suppose $y_0 = 10$. Define $y_1, y_2, y_3, \dots$ recursively via (2). Also define $y_{-1}, y_{-2}, y_{-3}, \dots$ recursively via the following equation for $t = 0, -1, -2, \dots$:

$$y_{t-1} = -\frac{\phi_0}{\phi_1} + \frac{y_t}{\phi_1} - \frac{\epsilon_t}{\phi_1} \tag{3}$$

   Note that (3) is just a restatement of (2) obtained by rearranging (2) with y_{t-1} on the left hand side. The resulting time series model will then clearly satisfy (2). However it will not be stationary because:

$$\mathrm{var}(y_0) = 0 \quad \text{and} \quad \mathrm{var}(y_1) = \mathrm{var}(\phi_0 + \phi_1 y_0 + \epsilon_1) = \mathrm{var}(\phi_0 + \epsilon_1) = \mathrm{var}(\epsilon_1) = \sigma^2.$$

2. Suppose $|\phi_1| < 1$ and define

$$y_t = \frac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}. \tag{4}$$

   The summation in the right hand side above is an infinite summation, and hence we need to address convergence issues. Because $|\phi_1| < 1$, the terms ϕ_1^j rapidly decay to 0 as j increases. This ensures that $\sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}$ is well-defined.

   It is easy to check that (4) satisfies (2) because:

$$\begin{aligned}
y_t &= \frac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j} \\
&= \frac{\phi_0}{1-\phi_1} + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \phi_1^3 \epsilon_{t-3} + \dots \\
&= \frac{\phi_0}{1-\phi_1} + \epsilon_t + \phi_1 \left(\epsilon_{t-1} + \phi_1 \epsilon_{t-2} + \phi_1^2 \epsilon_{t-3} + \dots\right) \\
&= \frac{\phi_0}{1-\phi_1} + \epsilon_t + \phi_1 \left(y_{t-1} - \frac{\phi_0}{1-\phi_1}\right) = \phi_0 + \phi_1 y_{t-1} + \epsilon_t.
\end{aligned}$$

   It is also true that (4) is a stationary model. This is because

$$\mathbb{E} y_t = \frac{\phi_0}{1-\phi_1} \qquad \text{for all } t$$

   and, for $h \geq 0$,

$$\begin{aligned}
\mathrm{cov}(y_t, y_{t+h}) &= \mathrm{cov}\left(\frac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}, \frac{\phi_0}{1-\phi_1} + \sum_{k=0}^{\infty} \phi_1^k \epsilon_{t+h-k}\right) \\
&= \mathrm{cov}\left(\sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}, \sum_{k=0}^{\infty} \phi_1^k \epsilon_{t+h-k}\right) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \phi_1^{j+k} \mathrm{cov}\left(\epsilon_{t-j}, \epsilon_{t+h-k}\right)
\end{aligned}$$

Because $\text{cov}\,(\epsilon_{t-j}, \epsilon_{t+h-k})$ is non-zero (equal to σ^2) only when $t - j = t + h - k$ i.e., $k = j + h$, we get

$$\text{cov}(y_t, y_{t+h}) = \sigma^2 \sum_{j=0}^{\infty} \phi_1^{2j+h} = \sigma^2 \frac{\phi_1^h}{1 - \phi_1^2}.$$

This clearly shows that $\{y_t\}$ is stationary with ACVF and ACF given by:

$$\gamma(h) = \sigma^2 \frac{\phi_1^{|h|}}{1 - \phi_1^2} \text{ and } \rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi_1^{|h|}.$$

We have thus proved that (4) is a stationary time series model that satisfies the AR(1) equation (2) when $|\phi_1| < 1$. In fact, it turns out that (4) is the only stationary solution of (2) when $|\phi_1| < 1$ (I am skipping proof of this). Thus (4) is the unique stationary AR(1) model when $|\phi_1| < 1$. The model (4) when $|\phi_1| < 1$ is known as the **causal stationary** AR(1) model. Causal refers to the fact that y_t is fully determined by present and past values of $\{\epsilon_t\}$.

3. Suppose $|\phi_1| > 1$ and define

$$y_t = \frac{\phi_0}{1 - \phi_1} - \sum_{j=1}^{\infty} \frac{\epsilon_{t+j}}{\phi_1^j}. \tag{5}$$

Note that y_t is well-defined because the infinite sum above has the coefficients ϕ_1^{-j} which decay rapidly as $|\phi_1| > 1$. It is easy to check that (5) also satisfies the AR(1) equation (2) and is stationary. In fact, it is the unique stationary AR(1) for $|\phi_1| > 1$. The model (5) is called the **non-causal, stationary** AR(1). It is non-causal because y_t depends on the future values of $\epsilon_{t+1}, \epsilon_{t+2}, \dots$.

For the model (5), it is certainly not true that ϵ_t is independent of $y_{t-1}, y_{t-2}, y_{t-3}, \dots$. Recall that we used this estimation while writing the likelihood for AR(1) for parameter estimation. Thus, if we attempt to estimate the parameters ϕ_0, ϕ_1, σ of (5) using our AR-parameter estimation technique, we will get incorrect an estimate of ϕ_1 (for more details, see Example 3.3 and 3.4 in the Shumway-Stoffer book 4th edition).

To summarize the above discussion, there exist many non-stationary AR(1) time series models. When $|\phi_1| < 1$, there exists a unique stationary AR(1) model that is given by the formula (4), this is called the causal, stationary AR(1) model. When $|\phi_1| > 1$, there also exists a unique stationary AR(1) model that is given by the formula (5), this is called the non-causal, stationary AR(1) model.

When $|\phi_1| = 1$ (i.e., $\phi_1 = 1$ or $\phi_1 = -1$), neither of the two formulae (4) and (5) are meaningful (i.e., the infinite series do not converge). Here it turns out that there is no stationary AR(1) model. To see this, consider the case $\phi_1 = 1$ (the case $\phi_1 = -1$ is similar) where

$$y_t = \phi_0 + y_{t-1} + \epsilon_t$$

This implies that for every $t \geq 1$

$$y_t - y_0 = t\phi_0 + \epsilon_1 + \dots + \epsilon_t$$

When $\phi_0 \neq 0$, clearly y_t and y_0 have different means (because $\mathbb{E}y_t = \mathbb{E}y_0 + t\phi_0$) so there is no stationarity. But even if $\phi_0 = 0$, we have

$$\text{var}(y_t - y_0) = \text{var}(\epsilon_1 + \dots + \epsilon_t) = t\sigma^2$$

which approaches ∞ as $t \uparrow \infty$. But if $\{y_t\}$ were stationary, we would have

$$\text{var}(y_t - y_0) \leq 2\text{var}(y_t) + 2\text{var}(y_0) \leq \text{constant}.$$

Thus there are two kinds of AR(1): stationary and non-stationary. Stationarity is only possible when $|\phi_1| \neq 1$. There are also two kinds of stationary AR(1) models. When $|\phi_1| < 1$, the stationary AR(1) model has the formula (4); this is the causal kind of stationarity. When $|\phi_1| > 1$, the stationary AR(1) model has the formula (5); this is the non-causal kind of stationarity.

References

[1] Shumway, R. H. and D. S. Stoffer (2010). *Time series analysis and its applications: with R examples* (fourth ed.). Springer Science & Business Media.