STAT 153 & 248 - Time Series
Lecture Twenty

Spring 2025, UC Berkeley

Aditya Guntuboyina

April 08, 2025

1 Time Series Models, and Stationarity

A time series model describes the distribution of random variables y_t for all t past and present i.e., $t = \dots, -3, -2, -1, 0, 1, 2, 3, \dots$, in terms of certain unknown parameters. We refer to this collection of random variables as doubly infinite because t extends to infinity in both directions.

All the time series models that we shall study in this course will be jointly Gaussian (i.e., the joint distribution of $(y_{t_1}, \dots, y_{t_k})$ will be multivariate Gaussian for every k and $t_1, \dots, t_k$). Gaussianity ensures that the behavior of the time series model is characterized by means (expectations) and covariances.

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

Definition 1.1 (Stationarity). *A doubly infinite sequence of random variables y_t 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:

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

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

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

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 two 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 (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:

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

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:

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

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.

2 Moving Average (MA) Models

MA models present the simplest examples of stationary time series models that are not just white noise. Given a positive integer $q \geq 1$, the Moving Average model with order q (denoted by MA(q)) is defined by the equation:

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{1}$$

where $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. The MA($q$) model has $q+2$ unknown parameters which are estimated from observed data: $\mu, \theta_1, \ldots, \theta_q, \sigma$.

The MA(q) model has been called the "Summation of Random Causes" by its inventor Slutzky in the original paper titled "The summation of random causes as the source of cyclic processes" published in Econometrica in 1937. Basically the ϵ_t's can be treated as random causes which are assumed to be independently and identically distributed. The actual observations y_t's are consequences of these causes. The consequence for tine t depends on the cause for time t as well as the causes for times $t-1, \ldots, t-q$. These different causes affect the consequence at time t differently depending on the values of $\theta_1, \ldots, \theta_q$. Note that successive observations y_t share some common causes leading to dependence between the successive values of y_t.

The mean of y_t is clearly equal to μ (so it is constant in t). The covariance between y_t and y_{t+h} is given by:

$$\begin{aligned}
&\operatorname{cov}(y_t, y_{t+h}) \\
&= \operatorname{cov}\left(\mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}, \mu + \epsilon_{t+h} + \theta_1 \epsilon_{t+h-1} + \theta_2 \epsilon_{t+h-2} + \cdots + \theta_q \epsilon_{t+h-q}\right) \\
&= \operatorname{cov}\left(\mu + \sum_{j=0}^{q} \theta_j \epsilon_{t-j}, \mu + \sum_{k=0}^{q} \theta_k \epsilon_{t+h-k}\right) \\
&= \sum_{j=0}^{q} \sum_{k=0}^{q} \theta_j \theta_k \operatorname{cov}\left(\epsilon_{t-j}, \epsilon_{t+h-k}\right).
\end{aligned}$$

Note that, in the sum $\sum_{j=0}^{q} \theta_j \epsilon_{t-j}$, we take $\theta_0 = 1$. Because $\{\epsilon_t\}$ is Gaussian white noise, the covariance $\operatorname{cov}(\epsilon_{t-j}, \epsilon_{t+h-k})$ equals zero unless $t-j = t+h-k$ i.e., $k = j+h$. So we need the three conditions $0 \leq j \leq q$, $0 \leq k \leq q$ as well as $k = j+h$. If $h > q$, it is clear that this is not possible for any j, k. So we have $\operatorname{cov}(y_t, y_{t+h})$ equals zero when $h > q$. When $0 \leq h \leq q$, we have $0 \leq j \leq q$ and $0 \leq j+h \leq q$ which implies $0 \leq j \leq q-h$. We then get

$$\operatorname{cov}(y_t, y_{t+h}) = \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}.$$

We thus have:

$$\operatorname{cov}(y_t, y_{t+h}) = \begin{cases} \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} & : 0 \leq h \leq q \\ 0 & : h > q \end{cases}$$

The above covariance does not depend on t which shows that the MA(q) model is stationary. Thus the ACVF of MA(q) is:

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} & : 0 \leq h \leq q \\ 0 & : h > q \end{cases}$$

The ACF $\rho(h) = \gamma(h)/\gamma(0)$ equals:

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{\sum_{j=0}^{q} \theta_j^2} & : 0 \leq h \leq q \\ 0 & : h > q \end{cases}$$

The simplest of these MA(q) models is MA(1) (i.e., $q = 1$):

$$y_t = \mu + \epsilon_t + \theta \epsilon_{t-1}.$$

The ACF of MA(1) is:

$$\rho(h) = \begin{cases} 1 & : h = 0 \\ \frac{\theta_1}{1+\theta_1^2} & : h = 1 \\ 0 & : h > 1 \end{cases}$$

3 Sample ACF

We have so far defined the ACF $\rho(h)$ for a stationary time series model as the correlation between y_t and y_{t+h}. Given a time series dataset $y_1, \ldots, y_n$, we can define a sample autocorrelation function that can be seen as an estimate of the ACF of a time series model that is assumed for the data.

For a fixed value of h, the sample acf at lag h is essentially defined as the correlation coefficient between $(a_t, b_t), t = 1, \ldots, n - h$ where $a_t = y_t$ and $b_t = y_{t+h}$. This correlation coefficient is given by:

$$\frac{\sum_{t=1}^{n-h} (a_t - \bar{a})(b_t - \bar{b})}{\sqrt{\sum_{t=1}^{n-h} (a_t - \bar{a})^2} \sqrt{\sum_{t=1}^{n-h} (b_t - \bar{b})^2}} = \frac{\sum_{t=1}^{n-h} (y_t - \bar{a})(y_{t+h} - \bar{b})}{\sqrt{\sum_{t=1}^{n-h} (y_t - \bar{a})^2} \sqrt{\sum_{t=1}^{n-h} (y_{t+h} - \bar{b})^2}}$$

where

$$\bar{a} = \frac{1}{n-h} \sum_{t=1}^{n-h} a_t = \frac{1}{n-h} \sum_{t=1}^{n-h} y_t \quad \text{and} \quad \bar{b} = \frac{1}{n-h} \sum_{t=1}^{n-h} b_t = \frac{1}{n-h} \sum_{t=1}^{n-h} y_{t+h}$$

This correlation can be simplified slightly by making the following approximations:

$$\bar{a} \approx \bar{y} \quad \bar{b} \approx \bar{y} \quad \sum_{t=1}^{n-h} (y_t - \bar{a})^2 \approx \sum_{t=1}^{n} (y_t - \bar{y})^2 \quad \text{and} \quad \sum_{t=1}^{n-h} (y_{t+h} - \bar{b})^2 \approx \sum_{t=1}^{n} (y_t - \bar{y})^2$$

which are reasonable when h is very small compared to n. Making these approximations lead to the following definition of the sample ACF:

$$r_h := \frac{\sum_{t=1}^{n-h} (y_t - \bar{y})(y_{t+h} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2} \qquad \text{for } h = 0, 1, 2, \ldots$$

Note that r_0 is always equal to 1.

The sample ACF $r_h, h \geq 0$ can be computed for any time series dataset, although it is only useful for data for which stationary models are appropriate.

The sample ACF is particularly useful for determining the order q for fitting an MA(q) model. For an MA(q) model, we have seen in the previous section that the theoretical ACF $\rho(h)$ becomes exactly zero when $h > q$. This suggests that if the sample ACF r_h for a particular dataset becomes small (not exactly zero because of randomness) when h exceeds a particular q, then MA(q) is probably a good model for that dataset. This diagnostic is very commonly used when working with MA models.

4 Sample PACF

As discussed above, the sample ACF is a useful diagnostic for determining the order of q for an MA(q) model given data. There exists a similar diagnostic called sample PACF (PACF stands for Partial AutoCorrelation Function) which is useful for determining the order of p for an AR(p) model.

The sample PACF is defined as follows: for $h \geq 1$,

$$\text{sample PACF}(h) = \text{estimate } \hat{\phi}_h \text{ of } \phi_h \text{ when AR}(h) \text{ is fit to the data}$$

If the sample PACF(h) becomes negligibly small after a particular p, this suggests that AR(p) is a good model for the data. This method is similar to the heuristic technique that we used in last lecture and Lab 10 for selecting the order p to fit AR(p). There we were looking at the uncertainty interval for ϕ_p to see if it contains zero when AR(p) is fit to the data. This is the same as checking whether the sample PACF(h) is small at $h = p$.

It can happen (we will see examples of this later) that the sample PACF(h) for $h = 1, 2, \ldots, 11$ are all negligible but at $h = 12$, it is nonnegligible. In that case, we would be using AR(12). More specifically, we will use that value of p for which sample PACF(h) is negligible for all $h > p$. The same is true for sample ACF and MA(q).

5 Why is this called "Partial" Autocorrelation?

Why should the quantity $\hat{\phi}_h$ (obtained by fitting AR(h) to the data) be called the Sample Partial Autocorrelation? It turns out that there is a connection between regression coefficients and something called "partial correlation" (see e.g., `https://en.wikipedia.org/wiki/Partial_correlation`).

Suppose we have data on two variables x and y: $(x_1, y_1), \ldots (x_n, y_n)$. The correlation between them is defined in the usual way as:

$$\text{corr}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Correlation is also related to regression. If we regress y_i on x_i and obtain the usual least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, then

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \text{corr}(x, y) \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = \text{corr}(x, y) \sqrt{\frac{\text{var}(y)}{\text{var}(x)}} \tag{2}$$

where $\text{var}(x) = \sum_{i=1}^n (x_i - \bar{x})^2$ and $\text{var}(y) = \sum_{i=1}^n (y_i - \bar{y})^2$. Note that, in this section, corr and var refer to things calculated on the observed data (as opposed to random variables).

Now instead of just having data on x and y, we also have data on other variables $z_1, \ldots, z_k$. The dataset now is $(y_i, x_i, z_{i1}, \ldots, z_{ik})$ for $i = 1, \ldots, n$. The partial correlation between x and y given the variables $z_1, \ldots, z_k$ is denoted by $\text{corr}(x, y \mid z_1, \ldots, z_k)$ and is defined as the correlation between the residual of x given $z_1, \ldots, z_k$ and the residual of y given $z_1, \ldots, z_k$.

Here, residual of x given $z_1, \ldots, z_k$ refers to the residual in the linear regression of x given $z_1, \ldots, z_k$:

$$e_i^{x|z_1,\ldots,z_k} := x_i - \hat{\beta}_0^x - \hat{\beta}_1^x z_{i1} - \cdots - \hat{\beta}_k^x z_{ik},$$

where $\hat{\beta}_0^x, \ldots, \hat{\beta}_k^x$ are the fitted regression coefficients of x_i on $1, z_{i1}, \ldots, z_{ik}$.

Similarly the residual in the linear regression of y given $z_1, \ldots, z_k$ is

$$e_i^{y|z_1,\ldots,z_k} := y_i - \hat{\beta}_0^y - \hat{\beta}_1^y z_{i1} - \cdots - \hat{\beta}_k^y z_{ik}$$

where $\hat{\beta}_0^y, \ldots, \hat{\beta}_k^y$ are the fitted regression coefficients of y_i on $1, z_{i1}, \ldots, z_{ik}$.

Therefore:

$$\operatorname{corr}(x, y \mid z_1, \ldots, z_k) = \operatorname{corr}\left(e^{x|z_1,\ldots,z_k}, e^{y|z_1,\ldots,z_k}\right).$$

Similar to (2), there is a nice relationship between fitted regression coefficients in multiple linear regression and partial correlation. Consider the multiple regression of y on x as well $z_1, \ldots, z_k$. Let the fitted regression coefficients be: $\hat{\beta}_0, \hat{\beta}_x, \hat{\beta}_1, \ldots, \hat{\beta}_k$:

$$(\hat{\beta}_0, \hat{\beta}_x, \hat{\beta}_1, \ldots, \hat{\beta}_k) \text{ minimize } \sum_{i=1}^{n} (y_i - \beta_0 - \beta_x x_i - \beta_1 z_{i1} - \cdots - \beta_k z_{ik})^2.$$

Then it turns out that

$$\hat{\beta}_x = \operatorname{corr}(x, y \mid z_1, \ldots, z_k)\sqrt{\frac{\operatorname{var}(e^{y|z_1,\ldots,z_k})}{\operatorname{var}(e^{x|z_1,\ldots,z_k})}}. \tag{3}$$

This is the connection between a fitted regression coefficient (corresponding to a specific covariate) in multiple linear regression and the partial correlation between the response and the covariate given the other covariates.

Now let us come to the time series setting with data $y_1, \ldots, y_n$. We fit AR(p) models using the regression:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t.$$

Following the formula (3), we write

$$\hat{\phi}_p = \operatorname{corr}(y_{t-p}, y_t \mid y_{t-1}, \ldots, y_{t-p+1})\sqrt{\frac{\operatorname{var}(e^{y_t|y_{t-1},\ldots,y_{t-p+1}})}{\operatorname{var}(e^{y_{t-p}|y_{t-1},\ldots,y_{t-p+1}})}}.$$

When the AR(p) model is stationary (we shall see in the next lecture on conditions for AR(p) models to be stationary), the population analogues of the variance terms above ($\operatorname{var}(e^{y_t|y_{t-1},\ldots,y_{t-p+1}})$ and $\operatorname{var}(e^{y_{t-p}|y_{t-1},\ldots,y_{t-p+1}})$) turn out to be equal, and they are nearly same in the sample so they can be dropped and we get

$$\hat{\phi}_p \approx \operatorname{corr}(y_{t-p}, y_t \mid y_{t-1}, \ldots, y_{t-p+1}).$$

$\operatorname{corr}(y_{t-p}, y_t \mid y_{t-1}, \ldots, y_{t-p+1})$ can be called the partial autocorrelation at lag p. This is the reason why the plot of $\hat{\phi}_h$ (when the AR(h) model is fit to the data) is referred to as the sample PACF plot.

6 Additional Optional Reading

1. For MA models, see parts of Section 3.2 of Shumway-Stoffer 4th edition.
2. For ACF and PACF, see Section 3.4 ofShumway-Stoffer 4th edition.