SF2943 Time Series Analysis: Lecture 2

March 31, 2022

In this lecture, we will cover more examples of stationary time series, including two very useful models in time series analysis called the moving-average process and the autoregressive process. After that, we will briefly explain the central ideas of traditional time series analysis. In particular, we will try to motivate why *stationary time series* is an important object to study. With this understanding, we will discuss more theoretical properties of a general stationary time series as well as study various important stationary time series models in the following lectures.

1 Examples of stationary time series: Continued

Example 1.1 (First-order moving average or MA(1) process) *Consider a time series* $\{X_t\}$ *defiend by*

$$X_t = Z_t + \theta Z_{t-1}, \quad t = 0, \pm 1, \pm 2, \dots$$

where $Z_t \sim WN(0, \sigma^2)$ *and* θ *is some real constant. Notice that* X_t *is sort of an "average" of* Z_{t-1} *and* Z_t *("moving" of* Z_{t-1} *for one step forward in time). We call such process the first-order moving average process. This process is stationary since its mean function* $\mu_X(t) = EX_t = EZ_t + \theta E_{t-1} = 0$*, so it is invariant with time. In addition,*

$$\begin{aligned}\gamma_X(t+h,t) &= Cov(X_{t+h}, X_t) = Cov(Z_{t+h} + \theta Z_{t+h-1}, Z_t + \theta Z_{t-1}) \\ &= Cov(Z_{t+h}, Z_t) + \theta Cov(Z_{t+h}, Z_{t-1}) + \theta Cov(Z_{t+h-1}, Z_t) + \theta^2 Cov(Z_{t+h-1}, Z_{t-1}).\end{aligned}$$

Now we need to proceed our derivation case by case. Observe that for $h \geq 2$*, the time points* $t, t-1, t+h, t+h-1$ *are all different. Since* $\{Z_t\}$ *is uncorrelated, we find* $\gamma_X(t+h,t) = 0$ *for* $h \geq 2$*. We can make a similar observation for* $h \leq -2$*. For* $h = 1$*,* $t+h-1$ *is the same as* t*, and other time points are all different. Therefore,* $\gamma_X(t+h,t) = \theta Cov(Z_{t+h-1}, Z_t) = \theta\sigma^2$ *for* $h = 1$*. This same applies to* $h = -1$*. Namely, we will have* $\gamma_X(t+h,t) = \theta Cov(Z_{t+h-1}, Z_t) = \theta\sigma^2$ *when* $h = -1$*. Lastly, when* $h = 0$*,*

$$\begin{aligned}\gamma_X(t+h,t) &= Cov(Z_t, Z_t) + \theta Cov(Z_t, Z_{t-1}) + \theta Cov(Z_{t-1}, Z_t) + \theta^2 Cov(Z_{t-1}, Z_{t-1}). \\ &= \sigma^2 + 0 + 0 + \theta^2\sigma^2 = (1+\theta^2)\sigma^2.\end{aligned}$$

To sum up, we find the covariance function

$$\gamma_X(t+h,t) = Cov(X_{t+h}, X_t) = \begin{cases} (1+\theta^2)\sigma^2, & \textit{if } h = 0 \\ \sigma^2, & \textit{if } h = 1, -1 \\ 0, & \textit{otherwise} \end{cases}$$

is invariant with time as well. This shows that $\{X_t\}$ *is stationary and its autocovariance function* $\gamma_X(h) = \gamma_X(t+h,t)$*. Moreover, we can compute its autocorrelation function*

$$\rho_X(h) \doteq \frac{\gamma_X(h)}{\gamma_X(h)} = \begin{cases} 1, & \text{if } h = 0 \\ \frac{\sigma^2}{(1+\theta^2)\sigma^2}, & \text{if } h = 1, -1 \\ 0, & \text{otherwise.} \end{cases}$$

Remark 1.1 *Intuitively speaking, since* X_t *is defined by an average of a while noise* Z *at two time* t *and* $t-1$*. This averaging "spreads out" the dependence to lag* $h = 1$ *and* -1*, i.e., makes* $\rho_X(1)$ *and* $\rho_X(-1)$ *non-zero. Recall that for* $Z_t \sim WN(0, \sigma^2)$*, we have* $\gamma_Z(h) = 0$ *for all* $h \neq 0$ *and* $\gamma_Z(0) = 1$*.*

The following example is more complicated. The existence of such a process is not apparent, so we will first assume the existence of such process, and assume this process is stationary. With these assumptions, we will try to calculate its mean function, acvf, and acf.

Example 1.2 (First-order autoregression or AR(1) process) *Assume that* $\{X_t\}$ *is a stationary time series satisfying*

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \pm 2, \ldots$$

where $Z_t \sim WN(0, \sigma^2)$*,* $|\phi| < 1$*, and assume that* Z_t *is uncorrelated with* X_s *for each* $s < t$*. We will try to calculate its mean function, acvf, and acf. First all, since we know* $\{X_t\}$ *is stationary, we denote* EX_t *by* μ*. Then we find that* $\mu = EX_t = \phi EX_{t-1} + EZ_t = \phi\mu + 0 = \phi\mu$*. This implies that* $EX_t = \mu = 0$ *due to* $|\phi| < 1$*. As for acvf* $\gamma_X(h)$*, we have to consider different* h*.*

For $h = 0$*,*

$$\begin{aligned} \gamma_X(0) &= Cov(X_t, X_t) = Cov(\phi X_{t-1} + Z_t, \phi X_{t-1} + Z_t) \\ &= \phi^2 Cov(X_{t-1}, X_{t-1}) + \phi Cov(X_{t-1}, Z_t) + \phi Cov(Z_t, X_{t-1}) + Cov(Z_t, Z_t). \\ &= \phi^2 \gamma_X(0) + 0 + 0 + \sigma^2 = \phi^2 \gamma_X(0) + \sigma^2, \end{aligned}$$

where the second last equation holds due to $\gamma_X(0) = Cov(X_t, X_t)$ *for any* t*, and we assume* Z_t *is uncorrelated with* X_s *for any time point* $s < t$*. Thus, we find*

$$\gamma_X(0) = \frac{\sigma^2}{1 - \phi^2}.$$

For $h \geq 1$*, a useful trick in the case is that we expand only the first term in the covariance,*

$$\begin{aligned} \gamma_X(h) &= Cov(X_{t+h}, X_t) = Cov(\phi X_{t+h-1} + Z_{t+h}, X_t) = \phi Cov(X_{t+(h-1)}, X_t) + Cov(Z_{t+h}, X_t). \\ &= \phi^2 \gamma_X(h-1) + 0 = \phi^2 \gamma_X(h-1) = \cdots = \phi^h \gamma_X(0) = \phi^h \frac{\sigma^2}{1 - \phi^2}. \end{aligned}$$

The fourth equation comes from the fact that Z_{t+h} *is uncorrelated with* X_s *for any time point* $s < t + h$ *and the definition of acvf* γ_X*.*

As for the case $h \leq -1$, we notice that for any h, since covariance is symmetric,

$$\gamma_X(h) = Cov(X_{t+h}, X_t) = Cov(X_t, X_{t+h}) = Cov(X_{(t+h)-h}, X_{t+h}) = \gamma_X(-h).$$

This shows that any acvf is an even function[1]*, and therefore, for $h \leq -1$*

$$\gamma_X(h) = \gamma_X(-h) = \phi^{-h}\frac{\sigma^2}{1-\phi^2}.$$

To summarize, the acvf is

$$\gamma_X(h) = Cov(X_{t+h}, X_t) = \begin{cases} \frac{\sigma^2}{1-\phi^2}, & \text{if } h = 0 \\ \phi^{|h|}\frac{\sigma^2}{1-\phi^2}, & \text{otherwise,} \end{cases}$$

and the acf is

$$\rho_X(h) \doteq \frac{\gamma_X(h)}{\gamma_X(h)} = \begin{cases} 1, & \text{if } h = 0 \\ \phi^{|h|}, & \text{otherwise.} \end{cases}$$

Remark 1.2 *At first glance, MA(1) process*

$$X_t = Z_t + \theta Z_{t-1}$$

and AR(1) process

$$X_t = \phi X_{t-1} + Z_t$$

look quite similar, but as we mentioned earlier, the X_t in MA(1) is an average of white noise Z at only two time points $t-1$ and t, so one expect that X_t is uncorrelated to X_{t+h} for $h \geq 2$. On the contrary, the X_t in AR(1) depends on X_{t-1}, which depends X_{t-2} and so on. Thus, one expects the "dependency" of X_t to other time points to last longer. This reflects that the acf $\rho_X(h)$ of AR(1), even though decays exponentially as $|h|$ increases, is nonzero for all h.

2 A central approach to time series modeling

In many areas of science, such as physics, if we wish to make predictions/forecasting, then we must assume that *something* does not vary with time. For example, if we observe that an object is moving towards a direction with a constant speed v, then we can predict the position of this object at time t by

$$x_t = x_0 + vt,$$

with x_0 the initial position. In a time series, this "invariance" assumption in our setting becomes requiring the mean function and covariance function to be invariant with time, which means that the time series is *stationary* according to our terminology. This also shows the importance of stationary time series as a fundamental object of analysis. However, how can we "dig out" the stationary time series given temporal data?

[1] A function f is an even function if for any x, we have $f(x) = f(-x)$.

2.1 Broad picture of time series analysis

In this subsection, we will provide a typical "recipe" to find the hidden stationary time series in our temporal data. We are not trying to make this introduction of recipe exhaustive, but rather to provide a broad and vague picture. See the textbook (section 1.4-1.6) for a more detailed introduction and comparison of various techniques.

2.1.1 Step1: Plot the data and make judgment

The first step is to plot the data out, and make our own judgment of the data. For instance, if we see an abrupt change in level, we might want to divide the time into smaller intervals and analyze them separately. Also, there might be a few data points that look like outliers, so we could consider removing them. Lastly, we have to decide if we can model our data by the "classical decomposition model" in time series analysis

$$X_t = m_t + s_t + Y_t.$$

This part requires some experience working on time series analysis, background knowledge about our data, and understanding of the "classical decomposition model". In this course, our focus is mainly on the case when we think the "classical decomposition model" works for our data. The other case might be addressed at the very end of the course.

2.1.2 Step2: Remove the trend and seasonal components

Assume that we decide to fit our data with the "classical decomposition model". Our next step is to remove the trend and seasonal components so that we will only have stationary time series left. The two primary removal approaches are 1. Estimation and 2. Elimination by differencing. A crucial difference between these two approaches is that the former aims at estimating the trend m_t as well the seasonal component s_t by techniques such as polynomial fitting, smoothing with a finite moving average, etc.; In contrast, the latter aims to eliminate them by differencing iteratively to produce a stationary time series. Both methods will end up with a stationary process (ideally).

2.1.3 Step 3: Decide a stationary model and use that for inference

Once we obtain a stationary time series, we have to decide which time series model to fit the data. As we see in our previous examples, since the avf of white noise, MA(1), and AR(1) all look pretty different, this makes avf a good tool for selecting the models. There could be other tools to distinguish different models if we study those models further. Therefore, in the next couple of lectures, we will introduce other stationary models and study their properties. We will also discuss stationary time series in general. This would be beneficial in understanding if a given data could be stationary or not. After we pick a model, we will be able to predict something in the future with our knowledge of this specific model.

2.2 Sample autocorrelation function

Although we have just said how to choose a model by computing the autocorrelation function, in practical problems, we do not start with a model but with *observed data* $\{x_1, x_2, \ldots, x_n\}$. Therefore, what we can do instead is to compute an estimation of the autocorrelation function, known as the sample autocorrelation function.

Definition 2.1 *Let $x_1, x_2, \ldots, x_n$ be observations of a time series. The sample mean of $x_1, x_2, \ldots, x_n$ is*

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

The sample autocovariance function (sample acvf) is

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}), \text{ for } -h < n < h.$$

The sample autocorrelation function (sample acf) is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \text{ for } -h < n < h.$$

Remark 2.2 *The form of the sample autocovariance function comes from the definition of the autocovariance function*

$$\gamma(h) = Cov(X_{t+h}, X_t) = E[(X_{t+h} - EX_{t+h})(X_t - EX_t)],$$

and since we estimate EX_t and EX_{t+h} by the sample mean $\bar{x}$.

Remark 2.3 *The reason that we divide the summation in the definition of sample acvf by $1/n$ instead of $1/(n - |h|)$ is to make $\hat{\rho}$ become a nonnegative definite function*[2]*. As we will see in the next section, nonnegative definite is an important property of a autocorrelation function, so we would like to make its estimation, sample autocorrelation function, also be nonnegative definite.*

Remark 2.4 *An observation of the summation appearing in the sample autocovariance function is that, given a lag h, the summation is summing over only $n - |h|$ terms instead of n terms. The reason is that for a data point such $x_{n-|h|+1}, \ldots, x_n$, there is no data point to pair with (since we don't have $x_{n+1}, \ldots, x_{n+h}$). This observation also indicates that for a larger lag h, we expect that $\hat{\gamma}(h)$ would be noisier (i.e., less accurate).*

Remark 2.5 *Notice that for any given data $x_1, \cdots, x_n$, we can always compute the sample acvf and sample acf. In fact, we could compute the sample acvf and sample acf, and then compare them with the properties of acvf and acf for time series to see how reasonable it is to assume our data comes from a time series.*

[2] The definition is given in the next section.

3 Stationary processes

In the previous section, we explain the importance of stationary processes in time series analysis. In the upcoming lectures, we will talk about the properties of stationary processes. We recall some definitions here: Given a time series $\{X_t\}$, the acvf of $\{X_t\}$ is

$$\gamma(h) = \mathrm{Cov}(X_{t+h}, X_t) \text{ for all } h \in \mathbb{Z}.$$

The acf of $\{X_t\}$ (with lag h) is

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

We omit the subscript X for simplicity.

Proposition 3.1 *Let $\{X_t\}$ be a stationary time series with acvf γ. Then γ satisfies*

- $\gamma(0) \geq 0$.
- $|\gamma(h)| \leq \gamma(0)$ *for all h.*
- γ *is a even function, i.e., $\gamma(h) = \gamma(-h)$ for all h.*

Proof. For the first one, by definition $\gamma(0) = \mathrm{Cov}(X_t, X_t) = \mathrm{Var}(X_t) \geq 0$. As for the second one, we recall the Cauchy-Schwartz inequality for covariance (this should be mentioned in a probability course): For any two random variables X, Y, we have

$$\mathrm{Cov}(X, Y) \leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}.$$

Hence,

$$|\gamma(h)| = |\mathrm{Cov}(X_{t+h}, X_t)| \leq \sqrt{\mathrm{Var}(X_{t+h})\mathrm{Var}(X_t)} = \gamma(0).$$

Lastly, as we have shown before, since covariance is symmetric,

$$\gamma_X(h) = \mathrm{Cov}(X_{t+h}, X_t) = \mathrm{Cov}(X_t, X_{t+h}) = \mathrm{Cov}(X_{(t+h)-h}, X_{t+h}) = \gamma_X(-h).$$

■

Another fundamental property of acvf is nonnegative definite.

Definition 3.2 *A function $\kappa : \mathbb{Z} \to \mathbb{R}$ is nonnegative definite (also known as positive semi-definite) if for any $n \in \mathbb{N}$, and for any $(a_1, \ldots, a_n) \in \mathbb{R}^n$, we have*

$$\sum_{i,j=1}^{n} a_i \kappa(i-j) a_j \geq 0.$$

Remark 3.3 *Equivalently, we can say a function $\kappa : \mathbb{Z} \to \mathbb{R}$ is nonnegative definite if for any $n \in \mathbb{N}$, the associated n-by-n matrix*

$$K_n = [\kappa(i-j)]_{i,j=1}^n = \begin{pmatrix} \kappa(0) & \kappa(1) & \cdots & \kappa(n-1) \\ \kappa(-1) & \kappa(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(-n+1) & \cdots & \cdots & \kappa(0) \end{pmatrix}$$

is a nonnegative definite matrix (in the sense defined in linear algebra). Namely, for any $\mathbf{a} = (a_1, \ldots, a_n)' \in \mathbb{R}^n$,

$$\mathbf{a}' K_n \mathbf{a} = \sum_{i,j=1}^{n} a_i (K_n)_{ij} a_j \geq 0.$$

This is equivalent to saying that every eigenvalue of K_n *is nonnegative.*

Theorem 3.4 (Characterization of Autocovariance Functions) *A function* $\gamma : \mathbb{Z} \to \mathbb{R}$ *is the autocovariance function of a stationary time series if and only if it is even and nonnegative definite.*[3]

[3] This is why we define the sample autocovariance functions in the way that they are even and nonnegative definite.