

# W271 Summer 2022 Lecture Video Question Solutions Week 6

## Contents

<b>Week 6 Time Series Analysis Part 1</b>	<b>1</b>
6.2 Introduction to Time Series Analysis . . . . .	1
6.3 Basic Terminology Used in Time Series Analysis . . . . .	2
6.9 Examining Time Series Dependency: Autocorrelation, Example 2 . . . . .	3
6.10 Notion of Stationarity: Definitions and Examples . . . . .	3

## Week 6 Time Series Analysis Part 1

### 6.2 Introduction to Time Series Analysis

**Q: Give two examples of the application of statistical modeling using time series data.**

**Solution:** There are numerous case studies of applying to statistical to time series data. Generally speaking, time series applications fall into a few categories:

- (1) Separating long term trends from seasonal and random shock terms to identify long range patterns; see time series decomposition models such as seasonal trend loess (STL)
- (2) Forecasting into the future using past values of the same series in univariate time series or other series in multivariate times series (this can also include other variables too beyond time series)
- (3) Testing hypotheses related to time series data such as a a time series has a specific seasonal structure or follows a specific pattern
- (4) Scenario planning / simulations which is related to forecasting; we can use a model of the time series data to forecast what may happen under different conditions

**Two examples of time series moeling are (1) predicting future stock prices and (2) predicting the unemployment rate.**

Lots of models to predict stock prices have a time series nature to them because there is autocorrelation in financial data i.e. observations near one another in time are correlated.

The same goes for the unemployment rate. Every quarter the Bureau of Labor Statistics releases information about the state of employment in the United States. Many economists and others use time series modeling to forecast what they think the unemployment will be.

## 6.3 Basic Terminology Used in Time Series Analysis

**Q: Write down an example of a discrete time stochastic process.**

**Solution:** Practically speaking a discrete time stochastic process means that (1) time is discrete and can be represented in discrete, integer periods such as  $t = 1, 2, 3, \dots$  and (2) the values of the random variable of interest  $X_t$  are related somehow to past values i.e.  $X_t = F_t(X_{t-1}, \dots, X_0)$ .

This relationship may be fully deterministic i.e.  $X_t$  is fully determined by its past values or involve some noise with a random component. The relationship can also change over time.

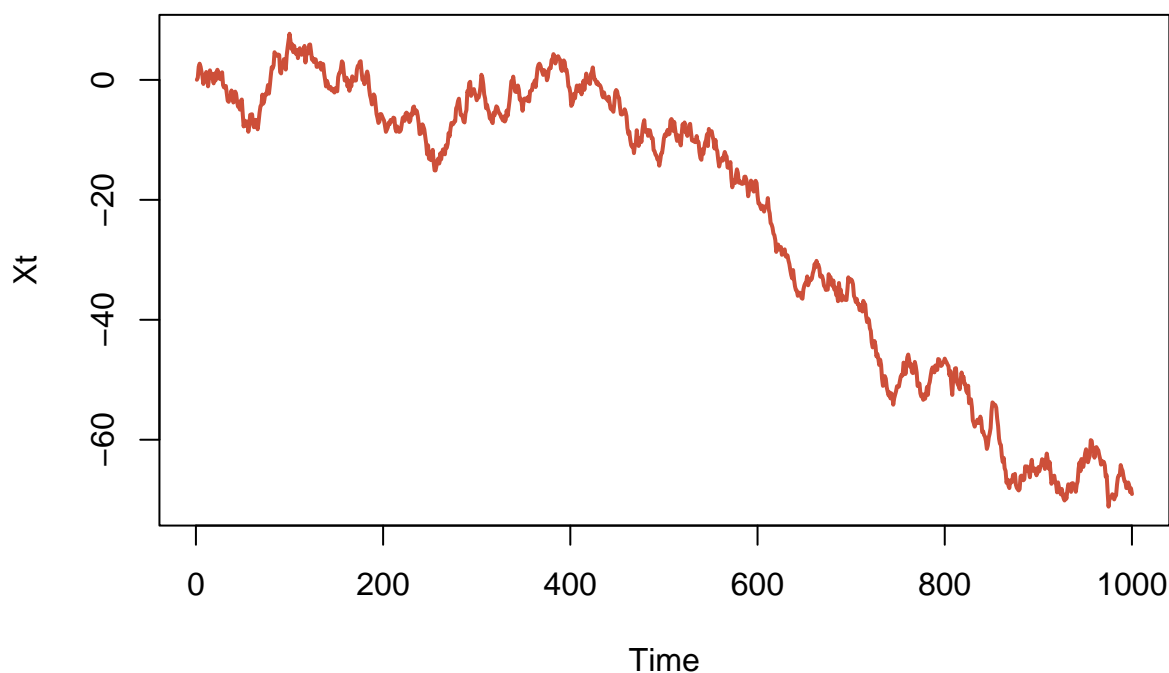
**One example of a discrete stochastic process that is very common in financial time series is the random walk:**

$$X_t = X_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0, \sigma^2)$$

Here  $X_t$  is equal to its past value plus some random noise. Here is an example of what it looks like:

```
X0 <- 0
sigma <- 1
n <- 1000
rw.time.series <- X0 + rnorm(n, 0, sigma)
rw.time.series <- cumsum(rw.time.series)

plot(rw.time.series, type = "l", xlab = "Time", ylab = "Xt", col = "tomato3", lwd = 2)
```



Visually and empirically this process matches a stock's price path quite well and is a central part of a lot of financial theory like the efficient market hypothesis.

## 6.9 Examining Time Series Dependency: Autocorrelation, Example 2

**Q:** Which time series model(s), if any, introduced so far is suitable for modeling the initial jobless claim series?

**Solution:** So far in the lectures a few basic time series models were discussed at a high level:

- (1) White noise i.e. time series is a series of uncorrelated random variables with no time dependence
- (2) Moving average model where the current period is a simple, equal weighted average of a fixed number of past periods
- (3) Autoregressive models where the current period is a linear function of the prior periods; note that the coefficients in this model are not time varying as it is a parametric, which makes it more rigid than the moving average model, but it does provide us with an explicit functional form
- (4) Random walk with drift where there is an overall trend in the time series

Jobless claims in the US generally speaking do not have a deterministic trend and exhibit a high degree of autocorrelation i.e. specific periods are highly related to several past periods. This is intuitive because structurally jobless claims should be related to the state of the economy and increases or decreases in the series in the past should generally be forecasted in with changes in the same direction in future periods.

**That means (1) and (4) do not match both the empirical trends in the times series and also the structural data generating process. But both (2) and (3) do capture the major trends of autocorrelation over time and a natural seasonal cycle.**

## 6.10 Notion of Stationarity: Definitions and Examples

**Q** Consider a random walk series that takes the following form where  $X_t = X_{t-1} + \omega_t$  and  $\omega_t$  is white noise. Answer the various questions listed.

**Solution:** To answer these questions, it helps to recursively substitute in the random walk form for past periods and define an initial period 1 and an arbitrary current period  $T$  i.e.:

$$X_T = \omega_T + X_{T-1} = \omega_T + \omega_{T-1} + X_{T-2} = \omega_T + \omega_{T-1} + \omega_{T-2} + X_{T-3} = \dots = \omega_T + \dots + \omega_1 = \sum_{t=1}^T \omega_t$$

**Write down the mean function of the process.**

Following the properties of white noise  $E(X_T) = E(\sum_{t=1}^T \omega_t) = \sum_{t=1}^T E(\omega_t) = 0$

So the mean function of a random walk is constant and  $E(X_t) = 0$ .

**Write down the variance function of the process.**

$Var(X_T) = Var(\sum_{t=1}^T \omega_t) = \sum_{t=1}^T Var(\omega_t)$  since  $\omega_t$  is independent across time as it is white noise

$Var(X_T) = \sum_{t=1}^T Var(\omega_t) = t\sigma^2$  since  $\omega_t$  is identically distributed across time

Therefore, the variance function of a random walk is  $Var(X_t) = t\sigma^2$ . Note this grows over time and approaches infinity as the time horizon gets very long, which is essentially why random walks despite being a sum of random noise tend to drift either completely up or down. There is a rich statistical literature on the interesting properties of the random walk. For an interesting use case see the gambler's ruin problem, which has applications to statistical testing as well.

**Write down the autocovariance function of the process.**

We will define the autocovariance function as the covariance between a period and its lag, which holds for any number of lags in the past, including zero.

$$Cov(X_T, X_{T-k}) = Cov(\sum_{t=1}^T \omega_t, \sum_{t=1}^{T-k} \omega_t) =$$

$Cov(\sum_{t=1}^{T-k} \omega_t, \sum_{t=1}^{T-k} \omega_t) = Var(\sum_{t=1}^{T-k} \omega_t)$  since  $\omega_t$  is independent across time as it is white noise, meaning  $Cov(\omega_{t_1}, \omega_{t_2}) = 0$  for  $t_1 \neq t_2$

$$Var(\sum_{t=1}^{T-k} \omega_t) = (T-k)\sigma^2$$

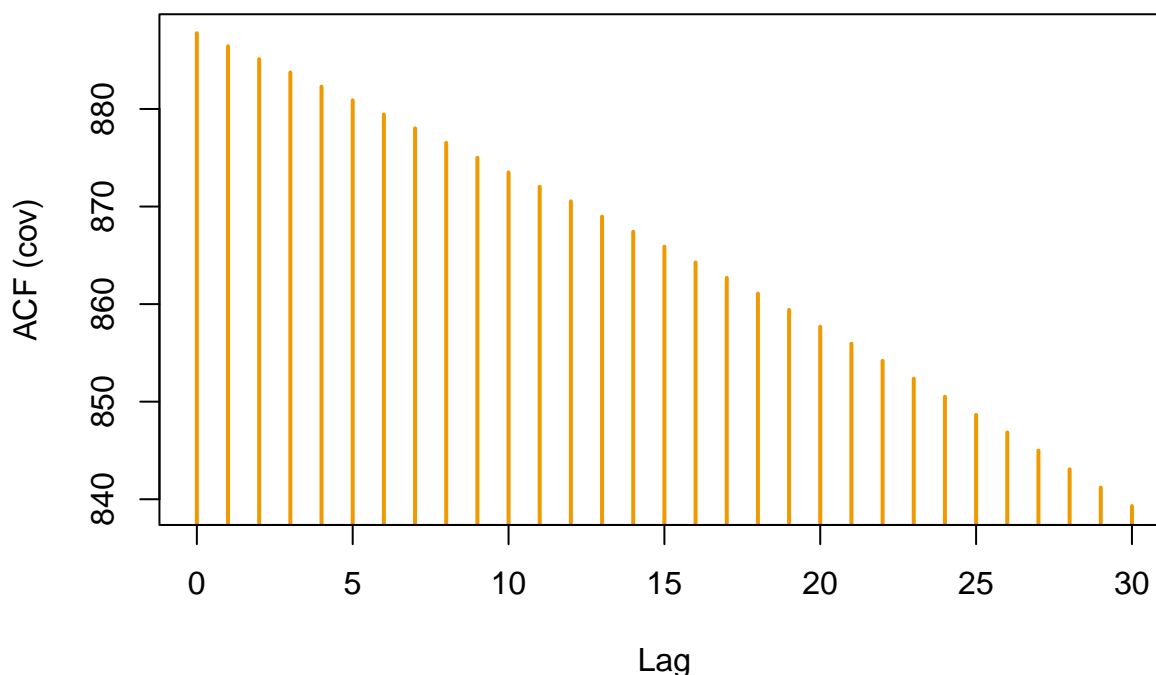
So the covariance function of the process is  $Cov(X_T, X_{T-k}) = (T-k)\sigma^2$  i.e. the covariance between the series decreases as the number of lags between periods increases (as a check the covariance with a lag of zero reduces to the variance like it should).

Empirically, it looks something like this:

```
X0 <- 0
sigma <- 1
n <- 1000
rw.time.series <- X0 + rnorm(n, 0, sigma)
rw.time.series <- cumsum(rw.time.series)

acf(rw.time.series, type = "covariance", lwd = 2, col = "orange2")
```

## Series rw.time.series



**Write down the autocorrelation function of the process.**

We will define the autocorrelation function as the correlation between a period and its lag, which holds for any number of lags in the past, including zero.

And remember that  $Corr(X, Y) = \frac{Cov(X, Y)}{SD(X)SD(Y)}$

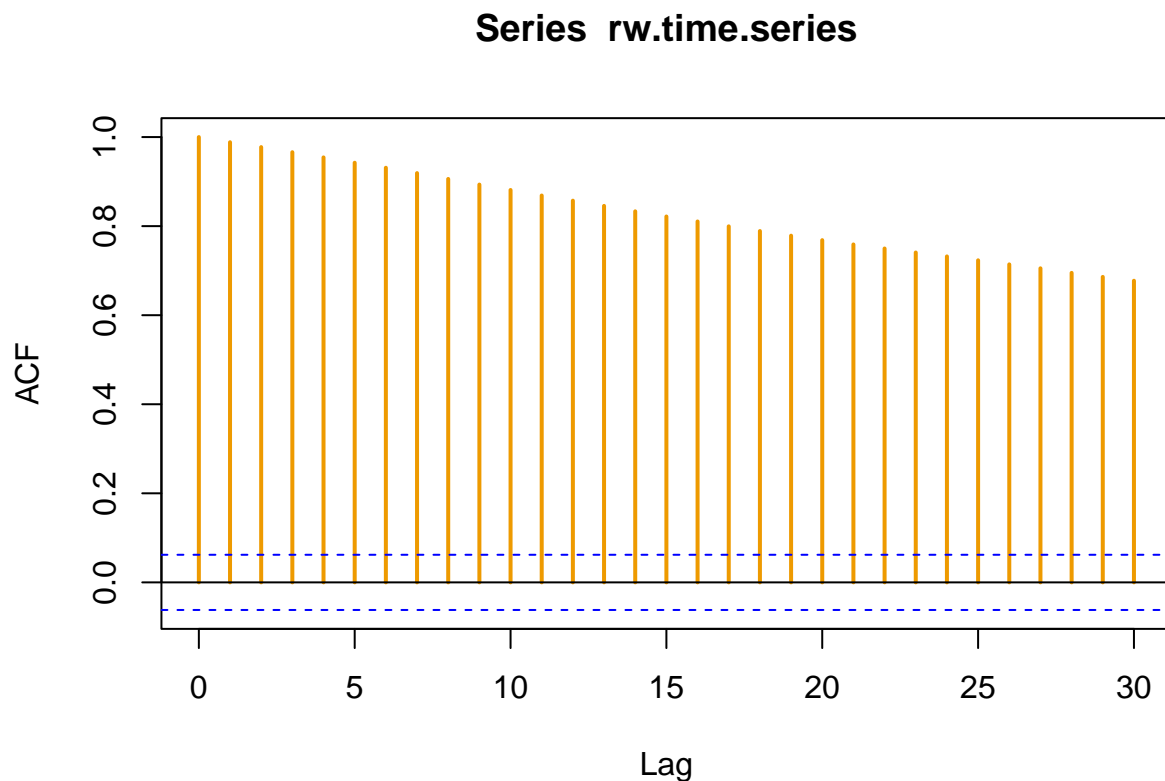
$$\begin{aligned}\text{So } \text{Corr}(X_T, X_{T-k}) &= \frac{\text{Cov}(X_T, X_{T-k})}{\sqrt{\text{Var}(X_T)\text{Var}(X_{T-k})}} = \frac{(T-k)\sigma^2}{\sqrt{T\sigma^2(T-k)\sigma^2}} \\ &= \frac{(T-k)\sigma^2}{\sigma^2\sqrt{T(T-k)}} = \sqrt{\frac{(T-k)^2}{T(T-k)}} = \sqrt{\frac{T-k}{T}} = \sqrt{1 - \frac{k}{T}}\end{aligned}$$

Therefore, the autocorrelation function  $\text{Corr}(X_T, X_{T-k}) = \sqrt{1 - \frac{k}{T}}$ , which again declines as the number of lags between periods increases in the autocorrelation function like the autocovariance function (as a check the correlation with a lag of zero reduces to one which is expected as period has a perfect correlation with itself).

Empirically, it looks something like this. Note how slowly it drops:

```
X0 <- 0
sigma <- 1
n <- 1000
rw.time.series <- X0 + rnorm(n, 0, sigma)
rw.time.series <- cumsum(rw.time.series)

acf(rw.time.series, type = "correlation", lwd = 2, col = "orange2")
```



### Is this process stationary?

There are various forms of stationarity in time series. The most common are strict / strong stationarity and weak stationarity.

Strict stationarity requires the cumulative distribution function of the joint distribution of  $X$  values to be constant over time. This means the data generating process of the time series does not change over time, which includes things like the mean, variance, autocovariance, and autocorrelation functions.

Weak stationarity requires that the mean and autocovariance function are stable over time and the variance of the series is finite for all time periods i.e. in the limit of time.

For the random walk, it is not strictly stationary since the variance function changes over time (it increases linearly with time). Note that the autocovariance and autocorrelation functions don't violate strict stationarity since the relationship does hold for fixed deltas between periods. It is also does not satisfy weak stationarity because the variance grows with time and therefore is not bounded below infinity for all time periods. Random walks tend to explode after a large number of time steps due to this fact.