A Glimpse at Time Series Analysis
Math 6070

Davar Khoshnevisan
University of Utah

April 20, 2014

Contents

1 Introduction **1**
- 1.1 The Autocorrelation Function 3
- 1.2 Cross-Covariance and Cross-Correlation 5
- 1.3 Stationarity 6

2 Exploratory Data Analysis **6**
- 2.1 Detrending 6
- 2.2 Differencing 7
- 2.3 Higher-Order Differencing 8
- 2.4 Other Useful Transformations of Data 10

3 Stationary Time Series **10**
- 3.1 Moment Analysis 10
- 3.2 Gaussian Time Series 14
- 3.3 Sampling Distribution of the Sample Mean 15
- 3.4 Sampling Distribution of the Sample ACF 19

1 Introduction

Time series are data that "arrive in time." In other words, a time series is a collection of random variables—or more appropriately put, a "stochastic process"—$\{x_t\}_{t=-\infty}^{\infty}$ that is indexed by a parameter t which we may think of as "time." It is more convenient to start time at $-\infty$ some times. Other times, one wants to start time at $t = 0$.

Example 1 (White noise)**.** A *white noise* $\{w_t\}_{t=-\infty}^{\infty}$ is an i.i.d. sequence of $\mathrm{N}(0\,,\sigma^2)$ random variables. I emphasize that, in this case, $\mathrm{E}(w_t) = 0$ and $\mathrm{Var}(w_t) = \sigma^2$ for all t. An important example is *Gaussian white noise*. That is the case where the w_t's are i.i.d. and each is distributed as $\mathrm{N}(0\,,\sigma^2)$. □

Example 2 (Moving Averages)**.** Let $w := \{w_t\}_{t=-\infty}^{\infty}$ be a white noise sequence. We can construct a new time series $x := \{x_t\}_{t=-\infty}^{\infty}$ from w as follows:

$$x_t := \mu + \frac{w_{t-2} + w_{t-1} + w_t}{3},$$

where μ is a fixed real number. This is an example of a *3-point moving-average model*. Note that, in this particular case,

$$\mathrm{E}(x_t) = \mu \quad \text{and} \quad \mathrm{Var}(x_t) = \frac{\sigma^2}{3},$$

for all times t. □

Example 3 (Autoregressive Models)**.** The simplest autoregressive model x, built from a white noise w, is a *2-point autoregressive model* that is defined as

$$x_s - x_{s-1} = \mu + w_s,$$

where μ is a fixed real number. We may add the preceding from $s = 1$ to $s = t$, say, and telescope the sum in order to see that

$$x_t - x_0 = \mu t + \sum_{t=1}^{t} (x_s - x_{s-1}) = \mu t + \sum_{s=1}^{t} w_s.$$

That is, $\{x_t - x_0\}_{t=0}^{\infty}$ is a *random walk* with drift μ. Here, $\mathrm{E}(x_t) = \mathrm{E}(x_0) + \mu t$ and $\mathrm{Var}(x_t) = t\sigma^2$.

For an example of a *2-point autoregressive model*, we may consider a stochastic model of the form $x_t = \mu + x_{t-1} + 2x_{t-2} + w_t$. This model has a unique and well-defined solution provided that x_0 and x_1 are well defined. □

Example 4 (Signal in Noise)**.** A typical example of such a model is a stochastic process $y := \{y_t\}$ that is defined, though an unknown signal x and white noise w, as follows: $y_t = \mu + x_t + w_t$. □

1.1 The Autocorrelation Function

If $\{x_t\}_{t=-\infty}^{\infty}$ is a time series, then its *autocovariance function* is

$$\gamma(s\,,t) := \mathrm{Cov}(x_s\,,x_t) = \mathrm{E}\left[(x_s - \mu(s))(x_t - \mu(t))\right],$$

where

$$\mu(t) := \mathrm{E}(x_t)$$

denotes the *mean function* of x. Clearly, γ is a symmetric function. That is,

$$\gamma(s\,,t) = \gamma(t\,,s).$$

Sometimes we may write μ_x and γ_x in order to emphasize that μ_x and γ_x are respectively the mean and autocovariance functions of the time series x.

The *autocorrelation function* [ACF] of x is

$$\rho(s\,,t) := \frac{\gamma(s\,,t)}{\sqrt{\gamma(s\,,s)\gamma(t\,,t)}}.$$

As we did with the mean and the autocovariance function, we may add a subscript x to ρ, as ρ_x, in order to emphasize that this ρ_x is indeed the ACF of x.

The autocovariance and the autocorrelation functions of x describe the dependence/correlation structure of the time series x. As such, we can learn about the time evolution of x from γ and/or ρ.

Example 5 (White noise)**.** If w is a white noise with variance σ^2, then

$$\gamma(s\,,t) = \mathrm{E}(w_s w_t) = \begin{cases} \sigma^2 & \text{if } s = t, \\ 0 & \text{if } s \neq t. \end{cases}$$

In other words, $\gamma(s\,,t) = \sigma^2 I\{s = t\}$, whence $\rho(s\,,t) = I\{s = t\}$. □

Example 6 (Moving Averages)**.** Let w denote a variance-σ^2 white noise and x the centered three-point moving average,

$$x_t := \frac{w_{t-2} + w_{t-1} + w_t}{3}.$$

Then, $\mu_t := \mathrm{E}(x_t) = 0$ [i.e., x is "centered"] and

$$\gamma(t\,,t) = \frac{1}{9}\mathrm{E}\left[(w_{t-2} + w_{t-1} + w_t)^2\right] = \frac{\sigma^2}{3}.$$

Also,

$$\gamma(t-1\,,t) = \frac{1}{9}\mathrm{E}\left[(w_{t-3}+w_{t-2}+w_{t-1})\,(w_{t-2}+w_{t-1}+w_t)\right] = \frac{2\sigma^2}{9},$$

and

$$\gamma(t-2\,,t) = \frac{1}{9}\mathrm{E}\left[(w_{t-4}+w_{t-3}+w_{t-2})\,(w_{t-2}+w_{t-1}+w_t)\right] = \frac{\sigma^2}{9}.$$

For all other pairs $(s\,,t)$, $\gamma(s\,,t)=0$. In such models, it is helpful to consider γ in terms of a new variable $|t-s|$—this is the socalled *lag* variable—as follows: $\gamma(s\,,t) = ({}^{\sigma^2}\!/9)\times\max\{0\,,3-|t-s|)\}$. In this case, we also have $\rho(s\,,t)=\gamma(s\,,t)/(\sigma^2/3) = ({}^1\!/3)\max\{0\,,3-|t-s|\}$. □

Example 7 (Autoregression)**.** Consider the autoregressive [random walk] example,

$$x_t = x_{t-1}+w_t \qquad (t\geqslant 1), \qquad x_0 = 0.$$

In this case, $x_t=\sum_{s=1}^t w_s$, and therefore,

$$\gamma(s\,,t) = \mathrm{E}\left(\sum_{r=1}^t w_r \times \sum_{u=1}^s w_u\right) = \mathrm{E}\left(\left|\sum_{r=1}^{\min(s,t)} w_r\right|^2\right) = \sigma^2\times\min(s\,,t).$$

Note that the preceding cannot be described in terms of lag alone. [This is an example of "non-stationarity."] Furthermore, $\rho(s\,,t)=\min(s\,,t)/\sqrt{st}$, which can be written as

$$\rho(s\,,t) = \sqrt{\frac{\min(s\,,t)}{\max(s\,,t)}},$$

for all times $s,t\geqslant 1$. □

Note that ρ is a unit-free function. The following shows the range of that function.

Proposition 8 (The Cauchy–Schwarz Inequality)**.** *For all s,t,*

$$|\gamma(s\,,t)| \leqslant \sqrt{\gamma(s\,,s)\cdot\gamma(t\,,t)}.$$

Equivalently, $-1\leqslant\rho(s\,,t)\leqslant 1$.

Proof. The inequality for the γ's is equivalent to $|\rho(s\,,t)| \leqslant 1$, which is the stated inequality for ρ. Therefore, it suffices to establish the inequality for the γ's only.

First, let us consider the case that $\gamma(s\,,s) = \mathrm{Var}(x_s) > 0$, and write

$$0 \leqslant \mathrm{Var}(ax_s - x_t) = a^2\gamma(s\,,s) + \gamma(t\,,t) - 2a\gamma(s\,,t) := f(a).$$

The minimum of f must therefore be nonnegative. That minimum can be found by using calculus: $f'(a) = 2a\gamma(s\,,s) - 2\gamma(s\,,t)$ and $f''(a) = 2\gamma(s\,,s) = \mathrm{var}(x_s) > 0$. Therefore, f is minimized at $a_{\min} = \gamma(s\,,t)/\gamma(s\,,s)$, and the minimum of f is

$$0 \leqslant f(a_{\min}) = \frac{|\gamma(s\,,t)|^2}{\gamma(s\,,s)} + \gamma(t\,,t) - 2\frac{|\gamma(s\,,t)|^2}{\gamma(s\,,s)} = \gamma(t\,,t) - \frac{|\gamma(s\,,t)|^2}{\gamma(s\,,s)}.$$

Solve to deduce the inequality, $|\gamma(s\,,t)|^2 \leqslant \gamma(s\,,s)\cdot\gamma(t\,,t)$; this is another way to state the Cauchy–Schwarz inequality. Because $\gamma(s\,,t) = \gamma(t\,,s)$, there is symmetry in the variables. In other words, the Cauchy–Schwarz inequality also holds when $\gamma(t\,,t) = \mathrm{Var}(x_t) > 0$. It remains to consider the case that $\gamma(s\,,s)$ and $\gamma(t\,,t)$ are both zero. But in that case, $x_s = \mathrm{E}(x_s)$ and $x_t = \mathrm{E}(x_t)$, whence we have $\gamma(s\,,t) = 0 = \sqrt{\gamma(s\,,s)\cdot\gamma(t\,,t)}$. □

1.2 Cross-Covariance and Cross-Correlation

If $x := \{x_t\}$ and $y := \{y_t\}$ are two time series, then their *cross covariance* is the function

$$\gamma_{x,y}(s\,,t) := \mathrm{Cov}(x_s\,,y_t) = \mathrm{E}\left[(x_s - \mu_x(s))\,(y_t - \mu_y(t))\right].$$

The corresponding *cross correlation* is the function

$$\rho_{x,y}(s\,,t) := \frac{\gamma_{x,y}(s\,,t)}{\sqrt{\gamma_x(s\,,s)\cdot\gamma_y(t\,,t)}}.$$

The following is a small generalization of the previous Cauchy–Schwarz inequality, and is proved by similar methods.

Proposition 9 (The Cauchy–Schwarz Inequality)**.** *The function $\rho_{x,y}$ is unit-free and takes values in $[-1\,,1]$.*

1.3 Stationarity

Definition 10. We say that $x := \{x_t\}$ is *weakly stationary* if $\gamma_x(s\,,t)$ depends only on the lag variable $|t-s|$. We say that x is *stationary* if the joint distribution of $(x_{t_1}\,,\ldots,x_{t_n})$ is the same as the joint distribution of $(x_{t_1+h}\,,\ldots,x_{t_n+h})$ for all $t_1,\ldots,t_n$ and h.

It is easy to see that stationarity implies weak stationarity. There are examples that show that the converse is not true.

Proposition 11 (Simple Properties of Stationary Time Series)**.** *If x is weakly stationary, then:*

1. *μ is a constant and $\gamma(0) = \mathrm{Var}(x_t)$ for all t;*
2. *$\gamma(s\,,t) = \gamma(0\,,|t-s|)$, therefore, we will frequently write $\gamma(t-s)$ in place of $\gamma(s\,,t)$. In this notation, $\gamma(h) = \gamma(-h)$ for all $h \in \mathbf{R}$;*
3. *$\rho(s\,,t) = \gamma(|t-s|)/\gamma(0)$ also depends only on the lag variable $|t-s|$. Therefore, we frequently write $\rho(t-s)$ instead, and note that $\rho(h) = \rho(-h)$ for all $h \in \mathbf{R}$.*

2 Exploratory Data Analysis

In general, we prefer to study stationary time series. For such time series, for example, one can use "averaging principles." What can be done if the data is not stationary? Two standard methods are "detrending" and "differencing."

2.1 Detrending

If the data exhibits "linear trends," then it cannot be stationary. In such cases, one can try to limit the effect on stationarity by fitting a straight line of the form $\beta_0 + \beta_1 t$ (for $t = 1,\ldots,n$, say) to the data, using $x_1,\ldots,x_n$. In other words, we posit the model,

$$x_t = \beta_0 + \beta_1 t + w_t,$$

where $\{w_t\}$ is white noise. In other words, we are supposing that the observable time series x is a linear perturbation of a time series w that is white, and hence has no linear trends. The preceding is a standard regression model and the least squares estimators of β_0 and β_1 are

$$\hat\beta_1 = \frac{\sum_{t=1}^n (x_t - \bar x)(t - \bar t)}{\sum_{t=1}^n (t-\bar t)^2} \quad \text{and} \quad \hat\beta_0 = \bar x - \hat\beta_1 \bar t,$$

where $\bar{x} := n^{-1} \sum_{t=1}^{n} x_t$ and $\bar{t} := n^{-1} \sum_{t=1}^{b} t = (n+1)/2$ the respective averages of $x_1, \ldots, x_n$ and $1, \ldots, n$. The "residuals,"

$$\hat{w}_t := x_t - \hat{\beta}_0 - \hat{\beta}_1 t$$

form a *detrended* series, and ought to be more or less free of linear trend.

2.2 Differencing

"Differencing" is an alternative EDA method to detrending. It is best to start with an example first.

Example 12 (A linear filter)**.** We can try the "linear filter,"

$$(\nabla x)_t := x_t - x_{t-1}.$$

The "operator" ∇ is called a "linear filter" because: (i) It is linear $[\nabla(ax + by) = a\nabla x + b\nabla y]$; and (ii) If we feed into ∇ a time series x, then we obtain a new time series ∇x.

The filter ∇ removes linear trends. Here is why: Suppose $x_t = \beta_0 + \beta_1 t + w_t$, where w is a stationary [resp. weakly stationary] series. Then,

$$(\nabla x)_t = \beta_1 + (\nabla w)_t$$

is also a stationary [resp. weakly stationary] series. □

To see how the linear filter ∇ works, consider atmospheric data by Uhse, Schmnidt, and Levin (`http://cdiac.ornl.gov/ftp/trends/co2/westerland.co2`). This time series is plotted in Figure 1 below.

The data describes a rather extensive time series of atmospheric CO_2 concentrations [in ppmv] that were gathered in Westerland, Germany during the years 1972–1997. In words, Figure 1 shows you a plot of all points of the form (t, x_t), where $x_t :=$ the concentration of CO_2 at time t. Now consider Figure 2 which is the plot of the linear filter of the previous example, applied to our atmospheric data.

Figure 2 shows you all points of the form $(t, (\nabla x)_t))$. Do you see how the filter ∇ removed the linear trends from the original time series? This suggests that the original atmospheric data has the form

$$x_t = \beta_0 + \beta_1 t + w_t,$$

where w is a stationary time series. [In fact, w is likely to be white noise in this case, but I have not run a test of independence.]

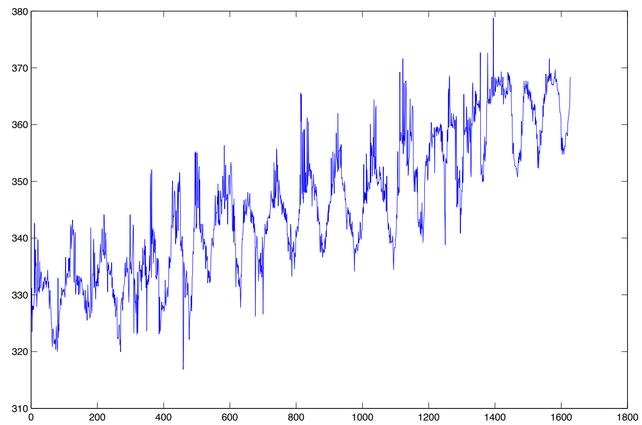

Figure 1: The actual CO_2 concentration data [ppmv against time]

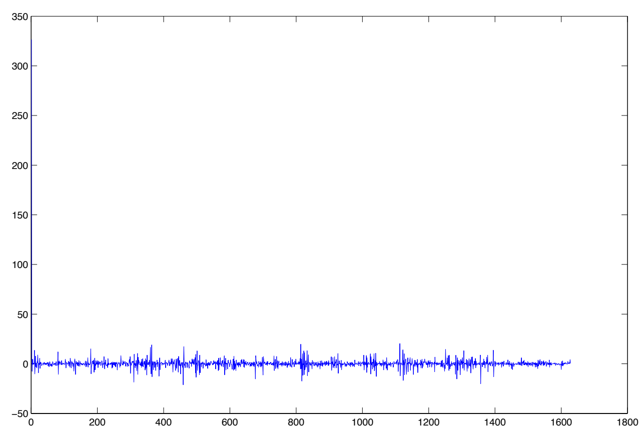

Figure 2: The CO_2 concentration data after an application of the difference filter ∇ [ppmv against time]

2.3 Higher-Order Differencing

Differencing methods can be used to also remove higher-order trends in time series.

Example 13. Consider the "quadratic filter" ∇^2 that is defined as follows:

$$\begin{aligned}(\nabla^2 x)_t &:= (\nabla(\nabla x))_t \\ &= (\nabla x)_t - (\nabla x)_{t-1} \\ &= (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) \\ &= x_t - 2x_{t-1} + x_{t-2}.\end{aligned}$$

We can also think of ∇^2 as the following "difference operator":

$$(\nabla^2 x)_t = 2\left(\frac{x_t + x_{t-2}}{2} - x_{t-1}\right).$$

Now suppose x is quadratic trends; i.e., that there exists a stationary series w [e.g., a white noise series] and numbers $\beta_0, \beta_1\, \beta_2$ such that

$$x_t = \beta_0 + \beta_1 t + \beta_2 t^2 + w_t.$$

In order to compute $\nabla^2 x$, let us first note that

$$\begin{aligned}(\nabla x)_t &= \left[\beta_0 + \beta_1 t + \beta_2 t^2 + w_t\right] - \left[\beta_0 + \beta_1(t-1) + \beta_2(t-1)^2 + w_{t-1}\right] \\ &= \beta_1 + \beta_2(2t-1) + (\nabla w)_t.\end{aligned}$$

Therefore,

$$\begin{aligned}(\nabla^2 x)_t &= [\beta_1 + \beta_2(2t-1) + (\nabla w)_t] - [\beta_1 + \beta_2(2(t-1)-1) + (\nabla w)_{t-1}] \\ &= 2\beta_2 + (\nabla w)_t - (\nabla w)_{t-1} \\ &= 2\beta_2 + (\nabla^2 w)_t.\end{aligned}$$

Since w is stationary, the series $\nabla^2 w$ is stationary too. Therefore, $\nabla^2 x$ is stationary also. □

Example 14. One can even apply higher-order difference operators, thanks to the recursive definition,

$$\left(\nabla^{k+1} x\right)_t := \left(\nabla\left(\nabla^k x\right)\right)_t.$$

Induction shows that whenever

$$x_t = \beta_0 + \beta_1 t + \cdots + \beta_{k+1} t^{k+1} + w_t,$$

then

$$\left(\nabla^{k+1} x\right)_t = (k+1)!\beta_{k+1} + \left(\nabla^{k+1} w\right)_t.$$

That is, ∇^{k+1} acts as a discrete $(k+1)$st derivative-type operator. [if you apply it to a $(k+1)$st degree polynomial, then you obtain $(k+1)!$ times the leading coefficient of that polynomial.] When w is, additionally, white noise, then $\nabla^{k+1} w$ is stationary, and therefore so is $\nabla^{k+1} x$. □

2.4 Other Useful Transformations of Data

Occasionally, people transform data in other useful ways, depending on the trends of the underlying time series. Two popular transformations are

$$y_t := \ln x_t, \quad \text{and} \quad z_t := \frac{x_t^\varepsilon - 1}{\varepsilon},$$

for $\varepsilon > 0$, when x is a positive time series. These transformations [nonlinear filters] remove the effect of large distributional tails of the x's: Even when x_t is large with reasonable probabilty, then y_t is typically not large. And $z_t \approx y_t$ for $\varepsilon \approx 0$. Indeed, $\lim_{\varepsilon \to 0} z_t = y_t$, because this assertion is another way to state that $\mathrm{d}x_t^\varepsilon / \mathrm{d}\varepsilon = \ln x_t$, which is an elementary fact from calculus.

3 Stationary Time Series

An important problem in time series analysis is to estimate efficiently the mean fucnction $\mu(t)$ and the autocorrelation function $\rho(s\,,t)$.

3.1 Moment Analysis

Recall that, when x is a stationary time series., $\mu := \mathrm{E}(x_t)$ does not depend on t. Therefore, it might be natural to have hopes for estimating μ, as long as we get the chance to observe $x_1, \ldots, x_n$ for a large enough time n. The most natural estimator of μ is, of course, the sample running average:

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

Proposition 15. *If x is stationary, then*

$$\mathrm{E}(\bar{x}_n) = \mu \quad \textit{and} \quad \mathrm{Var}(\bar{x}_n) = \frac{1}{n} \sum_{h=-n+1}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h).$$

Proof. Linearity of expectations shows that $\mathrm{E}(\bar{x}_n) = \mu$; that is, $\bar{x}_n$ is an unbiased estimator of μ. Next we compute the variance of $\bar{x}_n$.

Clearly,

$$\begin{aligned} \mathrm{Var}(\bar{x}_n) &= \frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} \mathrm{E}\left[(x_s - \mu)(x_t - \mu)\right] \\ &= \frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} \gamma(t - s). \end{aligned}$$

We split up the double sum according to the three cases that can arise: Either $t > s$; or $s > t$; or $s = t$. In other words, we write

$$\begin{aligned}
\mathrm{Var}(\bar{x}_n) &= \frac{1}{n^2}\sum_{s=1}^{n-1}\sum_{t=s+1}^{n}\gamma(t-s) + \frac{1}{n^2}\sum_{s=2}^{n}\sum_{t=1}^{s-1}\gamma(t-s) + \frac{1}{n^2}\sum_{s=1}^{n}\gamma(0) \\
&= \frac{2}{n^2}\sum_{s=1}^{n-1}\sum_{t=s+1}^{n}\gamma(t-s) + \frac{\gamma(0)}{n},
\end{aligned}$$

since $\gamma(t-s) = \gamma(s-t)$. Next we observe that

$$\begin{aligned}
\frac{2}{n^2}\sum_{s=1}^{n-1}\sum_{t=s+1}^{n}\gamma(t-s) &= \frac{2}{n^2}\sum_{s=1}^{n-1}\sum_{h=1}^{n-s}\gamma(h) \\
&= \frac{2}{n^2}\sum_{h=1}^{n-1}\sum_{s=1}^{n-h}\gamma(h) \\
&= \frac{2}{n^2}\sum_{h=1}^{n-1}(n-h)\gamma(h) \\
&= \frac{2}{n}\sum_{h=1}^{n-1}\left(1-\frac{h}{n}\right)\gamma(h).
\end{aligned}$$

Therefore,

$$\mathrm{Var}(\bar{x}_n) = \frac{2}{n}\sum_{h=1}^{n-1}\left(1-\frac{h}{n}\right)\gamma(h) + \frac{\gamma(0)}{n},$$

which is another way to state the proposition, since $\gamma(h) = \gamma(-h)$. □

The preceding result might suggest that different sorts of dependence structures can arise when we study time series:

1. If x is white noise, then $\gamma(u) = 0$ when $u \neq 0$; therefore, $\mathrm{Var}(\bar{x}_n) = 1/n \to 0$ as $n \to \infty$. This and Chebyshev's inequality together imply the following law of large numbers, which you know already:

$$\bar{x}_n \xrightarrow{\mathrm{P}} \mu \qquad \text{as } n \to \infty. \tag{1}$$

2. If

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty, \tag{2}$$

then

$$\mathrm{Var}(\bar{x}_n) \approx \frac{1}{n} \sum_{h=-\infty}^{\infty} \gamma(h), \tag{3}$$

and so (1) hold, as it did in the uncorrelated case. Condition (2) is called the condition of "short-range dependence." Note that the variance of $\bar{x}_n$ still goes to zero as constant over n. Therefore, the rate of convergence in (1) is as in the white noise case.

3. The analysis of "long-range dependence" case is quite a bit more involved. Suppose, for the sake of concreteness, that

$$\gamma(h) \approx \frac{C}{|h|^\alpha} \qquad \text{as } |h| \to \infty, \tag{4}$$

where $C, \alpha > 0$.[1] We are interested in the "long-range dependent" case; that is, when $\sum_{h=-\infty}^{\infty} |\gamma(h)| = \infty$. This means that $\alpha \leqslant 1$. Let us consider the case that $0 < \alpha < 1$. We will return to the case that $\alpha = 1$ in the next example. In this case, we have

$$\begin{aligned}
\mathrm{Var}(\bar{x}_n) &= \frac{2}{n} \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) \gamma(h) + \frac{\gamma(0)}{n} \\
&\approx \frac{2}{n} \sum_{h=K}^{n-1} \left(1 - \frac{h}{n}\right) \gamma(h) + \frac{\gamma(0)}{n} \qquad \text{as } n \to \infty,
\end{aligned}$$

where K is an arbitrary [but fixed] positive integer. If K is large enough, however, then $\gamma(h) \approx C/h^\alpha$ for all $h \geqslant K$. Therefore, for any such choice of K,

$$\begin{aligned}
\mathrm{Var}(\bar{x}_n) &\approx \frac{2C}{n} \sum_{h=K}^{n-1} \left(1 - \frac{h}{n}\right) h^{-\alpha} + \frac{\gamma(0)}{n} \\
&= \frac{2C}{n^\alpha} \cdot \frac{1}{n} \sum_{h=K}^{n-1} \left(1 - \frac{h}{n}\right) \left(\frac{h}{n}\right)^{-\alpha} + \frac{\gamma(0)}{n} \\
&\approx \frac{2C}{n^\alpha} \cdot \frac{1}{n} \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) \left(\frac{h}{n}\right)^{-\alpha} + \frac{\gamma(0)}{n}.
\end{aligned}$$

[1] To put this condition in somewhat more precise terms, we are assuming that $\lim_{|h|\to\infty} |h|^\alpha \gamma(h) = C$.

Now, $n^{-1}\sum_{h=1}^{n-1} G(h/n) \approx \int_0^1 G(y)\,\mathrm{d}y$ by the very definition of the Riemann integral. Apply this with $G(y) := (1-y)y^{-\alpha}$ in order to see that

$$\begin{aligned}\mathrm{Var}(\bar{x}_n) &\approx \frac{2C}{n^\alpha}\cdot\int_0^1 (1-y)y^{-\alpha}\,\mathrm{d}y + \frac{\gamma(0)}{n}\\ &\approx \frac{2C}{n^\alpha}\cdot\int_0^1 (1-y)y^{-\alpha}\,\mathrm{d}y,\end{aligned}$$

since $0<\alpha<1$. Recall that the *beta integral*

$$\mathrm{B}(a\,,b) := \int_0^1 y^{a-1}(1-y)^{b-1}\,\mathrm{d}y$$

satisfies

$$\mathrm{B}(a\,,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \qquad \text{where} \qquad \Gamma(\tau) := \int_0^\infty y^{\tau-1}\mathrm{e}^{-y}\,\mathrm{d}y$$

denotes the Gamma function. Thus,

$$\begin{aligned}\mathrm{Var}(\bar{x}_n) &\approx \frac{2C}{n^\alpha}\cdot \mathrm{B}(1-\alpha\,,2)\\ &= \frac{2C}{n^\alpha}\cdot\frac{\Gamma(1-\alpha)\Gamma(2)}{\Gamma(3-\alpha)}\\ &= \frac{2C\Gamma(1-\alpha)}{\Gamma(3-\alpha)}\cdot n^{-\alpha} \qquad \text{as } n\to\infty.\end{aligned}$$

So, once again, the consistency assertion (1) holds, but this convergence is slower than it was in the short-range dependence case.

4. Finally, let us consider the $\alpha = 1$ case. In that case, we still have $\sum_{h=-\infty}^{\infty} |\gamma(h)| = \infty$. However, the variance computations are different:

$$\begin{aligned}\mathrm{Var}(\bar{x}_n) &\approx \frac{2C}{n}\sum_{h=1}^{n-1}\left(1-\frac{h}{n}\right)\frac{1}{h} + \frac{\gamma(0)}{n}\\ &\approx \frac{2C}{n}\sum_{h=1}^{n-1}\frac{1}{h} + \frac{\gamma(0)}{n}\\ &\approx \frac{2C\ln n}{n} \qquad \text{as } n\to\infty.\end{aligned}$$

Once again, we have the law of large numbers (1), but convergence is faster than in the $\alpha<1$ case, though it is slower than in the short-range dependence case.

3.2 Gaussian Time Series

Suppose $x = \{x_t\}$ is a stationary time series that is also assumed to be Gaussian [or a Gaussian process, in the sense that we have seen already: For all $t_1, \ldots, t_n$, $(x_{t_1}, \ldots, x_{t_n})$ has a multivariate normal distribution].

Since x is a stationary Gaussian series, it follows that

$$\bar{x}_n \sim \mathrm{N}\left(0\,,\mathrm{Var}(\bar{x}_n)\right) = \mathrm{N}\left(0\,, \frac{1}{n}\sum_{h=-n+1}^{n-1}\left(1-\frac{|h|}{n}\right)\gamma(h)\right) \qquad \text{for all } n.$$

Therefore, if we had a nice sequence v_n and some constant $C > 0$ such that $\mathrm{Var}(\bar{x}_n) \approx Cv_n$, then we have the central limit theorem,

$$\frac{\bar{x}_n - \mu}{\sqrt{v_n}} \xrightarrow{d} \mathrm{N}\left(0\,, C\right) \qquad \text{as } n \to \infty.$$

Let us return to the preceding four examples, and use the preceding in order to study the convergence rates of the law of large numbers (1), in this Gaussian setting.

1. If x is Gaussian white noise with variance σ^2, then

$$\sqrt{n}\,(\bar{x}_n - \mu) \sim \mathrm{N}(0\,,\sigma^2) \qquad \text{for all } n.$$

2. In the short-range dependent case where $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, we have seen that $\mathrm{Var}(\bar{x}_n) \approx n^{-1}\sum_{h=-\infty}^{\infty}\gamma(h)$. Therefore,

$$\sqrt{n}\,(\bar{x}_n - \mu) \xrightarrow{d} \mathrm{N}\left(0\ ,\ \sum_{h=-\infty}^{\infty}\gamma(h)\right) \qquad \text{as } n \to \infty.$$

3. In the long-range dependent setting where $\gamma(h) \approx C/|h|^\alpha$ as $|h| \to \infty$, where $0 < \alpha < 1$, then

$$n^{\alpha/2}\,(\bar{x}_n - \mu) \xrightarrow{d} \mathrm{N}\left(0\ ,\ \frac{\Gamma(3-\alpha)}{2C\Gamma(1-\alpha)}\right) \qquad \text{as } n \to \infty.$$

4. Finally, in the long-range dependent setting where $\gamma(h) \approx C/|h|$ as $|h| \to \infty$, we have

$$\sqrt{\frac{n}{\ln n}}\,(\bar{x}_n - \mu) \xrightarrow{d} \mathrm{N}\left(0\ ,\ \frac{1}{2C}\right) \qquad \text{as } n \to \infty.$$

3.3 Sampling Distribution of the Sample Mean

There are instances where x is a non-Gaussian time series, and yet one can still establish the asymptotic normality of $\bar{x}_n$. Here, we study a special case of a well-known result of this general type.

Let us consider a *linear process* of the form,

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \Psi_j w_{t-j},$$

where w denotes a white noise sequence with variance σ^2, and the Ψ_j's form a non-random sequence that is absolutely summable; that is,

$$\mathscr{S} := \sum_{j=-\infty}^{\infty} |\Psi_j| < \infty.$$

Since $\mu_x(t) = \mu$ for all t, it follows readily that

$$\begin{aligned}\gamma_x(s\,,t) &= \mathrm{E}\left[(x_t - \mu_x(t))(x_s - \mu_x(s))\right] \\ &= \mathrm{E}\left(\sum_{j=-\infty}^{\infty} \Psi_j w_{t-j} \cdot \sum_{i=-\infty}^{\infty} \Psi_i w_{s-i}\right) \\ &= \sum_{j=-\infty}^{\infty}\sum_{i=-\infty}^{\infty} \Psi_i \Psi_j \gamma_w(t-j\,,s-i).\end{aligned}$$

But $\gamma_w(t-j\,,s-i) = 0$ unless $t-j = s-i$, in which case $\gamma_w(t-j\,,s-i) = \sigma^2$. Because $t-j = s-i$ if and only if $i = j-(t-s)$, it follows that

$$\gamma_x(s\,,t) = \sigma^2 \sum_{j=-\infty}^{\infty} \Psi_{j-(t-s)}\Psi_j = \sigma^2 \sum_{j=-\infty}^{\infty} \Psi_{j-(t-s)}\Psi_j.$$

Note that the preceding shows that x is weakly stationary, and therefore,

$$\gamma(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \Psi_{j+h}\Psi_j \qquad \text{for all } h = 0, \pm 1, \pm 2, \ldots. \tag{5}$$

Lemma 16. *x is short-range dependent, therefore, $\bar{x}_n \overset{\mathrm{P}}{\to} \mu$ as $n \to \infty$. Moreover,*

$$\sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \Psi_j\right)^2.$$

Proof. We apply the triangle inequality for sums in order to deduce the following:

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| \leqslant \sigma^2 \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |\Psi_{j+h}| \cdot |\Psi_j|$$
$$= \sigma^2 \left(\sum_{k=-\infty}^{\infty} |\Psi_k| \right)^2 = \sigma^2 \mathscr{S}^2 < \infty.$$

This proves short-range dependence, and the law of large numbers ensues, since we have seen already that $\mathrm{Var}(\bar{x}_n) \approx C/n$ in the short-range dependence case; see (3) above. The formula for $\sum_{h=-\infty}^{\infty} \gamma(h)$ is derived by going through the displayed computation above. □

Theorem 17 (Asymptotic Normality)**.** *If* $\sum_{j=-\infty}^{\infty} \Psi_j \neq 0$, *then*

$$\sqrt{n}\,(\bar{x}_n - \mu) \xrightarrow{d} \mathrm{N}\left(0\ ,\ \sigma^2 \left[\sum_{j=-\infty}^{\infty} \Psi_j \right]^2 \right).$$

Sketch of Proof. For every integer $N \geqslant 1$, define a new time series

$$x_t^N := \mu + \sum_{j=-N}^{N} \Psi_j w_{t-j}.$$

Note that x^N is itself a linear process; it is defined as x was, but instead of Ψ_j, we use $\Psi_j I\{|j| \leqslant N\}$. Moreover, $x - x^N$ is also a linear process, but we use $\mu \equiv 0$ and Ψ_j gets replaced by $\tilde{\Psi}_j := \Psi_j I\{|j| > N\}$. Since $\sum_{j=-\infty}^{\infty} |\tilde{\Psi}_j| = \sum_{|j|>N} |\Psi_j| \leqslant \sum_{j=-\infty}^{\infty} |\Psi_j| < \infty$, we may apply (5) in this case as well, and find that for every $h := t - s = 0, \pm 1, \pm 2 \ldots,$

$$\mathrm{Cov}\left(x_t - x_t^N, x_s - x_s^N\right) = \sigma^2 \sum_{j=-\infty}^{\infty} \tilde{\Psi}_{j+(t-s)} \tilde{\Psi}_j.$$

We may therefore average from $t = 1$ to $t = n$ and $s = 1$ to $s = n$ in order to find that

$$\begin{aligned}
\mathrm{E}\left(\left|\bar{x}_n - \bar{x}_n^N\right|^2\right) &= \mathrm{Var}\left(\bar{x}_n - \bar{x}_n^N\right) \\
&= \frac{1}{n^2}\sum_{t=1}^n\sum_{s=1}^n \mathrm{Cov}\left(x_t - x_t^N\,, x_s - x_s^N\right) \\
&= \frac{\sigma^2}{n^2}\sum_{t=1}^n\sum_{s=1}^n\sum_{j=-\infty}^{\infty} \tilde{\Psi}_{j+(t-s)}\tilde{\Psi}_j \\
&\leqslant \frac{\sigma^2}{n^2}\sum_{t=1}^n\sum_{s=1}^n\sum_{|j|>N} |\Psi_{j+(t-s)}|\cdot|\Psi_j|.
\end{aligned}$$

In order to understand the behavior of this triple sum we rearrange the sum by first adding over s and t, and then adding over j. Here is the first computation that we will need in order to carry out this program: For every integer j,

$$\begin{aligned}
\sum_{t=1}^n\sum_{s=1}^n |\Psi_{j+(t-s)}| &= \sum_{t=1}^n\sum_{k=j+t-1}^{j+t-n} |\Psi_k| \\
&\leqslant n\sum_{k=-\infty}^{\infty}|\Psi_k| = n\mathscr{S}.
\end{aligned}$$

Therefore,

$$\mathrm{E}\left(\left|\bar{x}_n - \bar{x}_n^N\right|^2\right) \leqslant \frac{\sigma^2\mathscr{S}}{n}\sum_{|j|>N}|\Psi_j|.$$

In other words,

$$\lim_{n\to\infty}\mathrm{E}\left(\left|\sqrt{n}\left[\bar{x}_n - \bar{x}_n^N\right]\right|^2\right) \leqslant \sigma^2\mathscr{S}^2\cdot\sum_{|j|>N}|\Psi_j|,$$

which can be made to be as small as we wish, since N can be made to be very very large, even though it is fixed and does not go to ∞ with n. It turns out that such an approximation reduces the problem to one about showing that $\bar{x}_n^N$ is asymptotically normal.

Now, it it true that $\bar{x}_n^N$ is not the average of n independent random variables. But it is the average of "$2N$-dependent" random variables in the following sense: Because

$$x_t^N = \mu + \sum_{|j|\leqslant N}\Psi_j w_{t-j}$$

depends only on $w_{t-N}, \ldots, w_{t+N}$, it follows that x_t and x_s are independent whenever $|t-s| > 2N$. The remainder of the theorem follows fairly readily from an application of a central limit theorem for $2N$-dependent sequences. I will state that CLT next, and sketch its proof. $\square$

We say that $X_1, X_2, \ldots$ is an *L-dependent sequence* when $(X_1, \ldots, X_L)$, $(X_{L+1}, \cdots, X_{2L})$, $(X_{2L+1}, \ldots, X_{3L})$, ... are i.i.d. L-dimensional random vectors.

Theorem 18 (CLT for L-dependent sequences)**.** *Suppose $X_1, X_2, \ldots$ is a stationary L-dependent sequence with* $\mathrm{E}(X_i) = 0$, $\mathrm{E}(X_i^2) = 1$, *and*

$$\mathrm{Cov}(X_i\,, X_j) = \begin{cases} \gamma(|j-i|) & \text{if } |j-i| \leqslant L, \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}\left(0\ ,\ \sum_{u=-L}^{L} \gamma(u)\right) \qquad \text{as } n \to \infty.$$

Sketch of Proof. The idea is to "block." If n and m are two positive integers such that $n \gg m \gg L$, then we write

$$X_1 + \cdots + X_{nm} := Z_1 + \cdots + Z_n,$$

where

$$\begin{aligned} Z_1 &:= X_1 + \cdots + X_m, \\ Z_2 &:= X_{m+1} + \cdots + X_{2m}, \\ &\vdots \\ Z_n &:= X_{(n-1)m+1} + \cdots + X_{nm}. \end{aligned}$$

Since X is L-dependent and stationary, and because $m > L$, it follows that the Z_i's are i.i.d. Moreover,

$$\begin{aligned} \mathrm{E}(Z_1) &= 0, \\ \mathrm{Var}(X_1) &= \mathrm{Var}\left(X_1 + \cdots + X_m\right) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{m} \gamma(i-j) \\ &= \sum_{u=-m}^{m} \left(m - |u|\right) \gamma(u), \end{aligned}$$

as we have seen before. The classical CLT can be applied to $Z_1, \ldots, Z_n$ in order to deduce that

$$\frac{X_1 + \cdots + X_{nm}}{\sqrt{n}} \xrightarrow{d} \mathrm{N}\left(0\,,\ \sum_{u=-m}^{m} (m - |u|)\,\gamma(m)\right) \qquad \text{as } n \to \infty.$$

Equivalently,

$$\frac{X_1 + \cdots + X_{nm}}{\sqrt{nm}} \xrightarrow{d} \mathrm{N}\left(0\,,\ \sum_{u=-m}^{m} \left(1 - \frac{|u|}{m}\right)\gamma(u)\right) \qquad \text{as } n \to \infty.$$

So far, I have shown you all of the steps of the proof completely. Now we have to rescale by setting $k := nm$ and then finally sending $m \to \infty$ [this part needs some care] in order to see that

$$\frac{X_1 + \cdots + X_k}{\sqrt{k}} \xrightarrow{d} \mathrm{N}\left(0\,,\ \sum_{u=-\infty}^{\infty} \gamma(u)\right) \qquad \text{as } n \to \infty.$$

This will prove the result, since $\sum_{u=-\infty}^{\infty} \gamma(u) = \sum_{u=-L}^{L} \gamma(u)$, by the very construction of γ. □

3.4 Sampling Distribution of the Sample ACF

Let us continue to examine the linear process of the previous subsection, subject to the short-range dependency condition that $\mathscr{S} < \infty$. Since x is a stationary time series, we might expect to estimate $\gamma(h)$ by the *sample autocovariance function*,

$$\hat{\gamma}_n(h) := \frac{1}{n}\sum_{t=1}^{n}(x_{t+h} - \bar{x}_n)(x_t - \bar{x}_n).$$

Theorem 19. *Suppose* $\mathrm{E}(|w_0|^4) := \eta\sigma^4 < \infty$ *and* $\sum_{j=-\infty}^{\infty} \Psi_j \neq 0$. *Then for each* $h = 0, \pm 1, \pm 2, \cdots$,

$$\sqrt{n}\,(\hat{\gamma}_n(h) - \gamma(h)) \xrightarrow{d} \mathrm{N}\left(0\,, (\eta - 3)\sigma^4 |\gamma(h)|^2 + \sum_{k=-\infty}^{\infty} \upsilon_k\right),$$

as $n \to \infty$, *where* $\upsilon_k := |\gamma(k)|^2 + \gamma(k+h)\gamma(k-h)$.

The proof is difficult, and requires laborious computations. Therefore, I will only sketch only enough of the ideas behind the proof so that you can see why the preceding theorem is true.

We wish to analyze the estimator $\hat{\gamma}_n(h)$, but as it turns out, it is helpful to introduce another random sequence first. Recall that $\mu(t) = \mu$ does not depend on time because x is stationary. Therefore, we may define

$$\tilde{\gamma}_n(h) := \frac{1}{n}\sum_{t=1}^{n}(x_{t+h} - \mu)(x_t - \mu).$$

The following shows that many of the interesting asymptotic properties of $\hat{\gamma}$ and $\tilde{\gamma}$ are equivalent, and reduces the proof of Theorem 19 to one about the asymptotic normality of $\tilde{\gamma}_n(h)$.

Lemma 20. *For every $h = 0, \pm 1, \pm 2, \ldots$,*

$$\sqrt{n}\left(\hat{\gamma}_n(h) - \tilde{\gamma}_n(h)\right) \xrightarrow{\mathrm{P}} 0 \quad \textit{as } n \to \infty.$$

Proof. We simply expand both quantities:

$$\begin{aligned}\hat{\gamma}_n(h) &= \frac{1}{n}\sum_{t=1}^{n}\left(x_{t+h}x_t - \bar{x}_n\left(x_{t+h} + x_t\right) + (\bar{x}_n)^2\right)\\ &= \frac{1}{n}\sum_{t=1}^{n} x_{t+h}x_t - \frac{\bar{x}_n}{n}\sum_{t=1}^{n} x_{t+h},\end{aligned}$$

and

$$\begin{aligned}\tilde{\gamma}_n(h)&\frac{1}{n}\sum_{t=1}^{n}\left(x_{t+h}x_t - \mu\left(x_{t+h} + x_t\right) + \mu^2\right)\\ &= \frac{1}{n}\sum_{t=1}^{n} x_{t+h}x_t - \frac{\mu}{n}\sum_{t=1}^{n} x_{t+h} - \mu\bar{x}_n + \mu^2.\end{aligned}$$

Consequently,

$$\begin{aligned}\hat{\gamma}_n(h) - \tilde{\gamma}_n(h) &= -(\bar{x}_n - \mu)\cdot\frac{1}{n}\sum_{t=1}^{n} x_{t+h} - \mu\left(\bar{x}_n - \mu\right)\\ &= -\left(\bar{x}_n - \mu\right)\left[\frac{1}{n}\sum_{t=1}^{n}(x_{t+h} - \mu)\right].\end{aligned}$$

We have seen already that $\sqrt{n}(\bar{x}_n - \mu)$ converges in distribution, as $n \to \infty$, to $\mathrm{N}(0\,, (\sum_j \Psi_j)^2)$. Also, $n^{-1}\sum_{t=1}^{n}(x_{t+h} - \mu)$ converges to zero in probability, as a result of short-range dependence. Therefore, the lemma follows from Slutsky's theorem. □

Now we analyze $\tilde{\gamma}_n(h)$, in hopes that it is asymptotically normal. First, we may note that $\tilde{\gamma}_n(h)$ is an unbiased estimator of $\gamma(h)$:

$$\mathrm{E}\left(\tilde{\gamma}_n(h)\right) = \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left[(x_{t+h}-\mu)(x_t-\mu)\right] = \gamma(h).$$

Next, we write

$$|\tilde{\gamma}_n(h)|^2 = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}(x_{t+h}-\mu)(x_{s+h}-\mu)(x_t-\mu)(x_s-\mu),$$

and use the relation $x_r - \mu = \sum_j \Psi_j w_{r-j}$ to simplify the preceding expression to the following:

$$\begin{aligned}&|\tilde{\gamma}_n(h)|^2\\ &= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{i-=\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty}\Psi_i\Psi_j\Psi_k\Psi_\ell w_{t+h-i}w_{s+h-j}w_{t-k}w_{s-\ell}.\end{aligned}$$

Then we compute directly the expectation of the preceding, in terms of the quantities

$$\mathrm{E}\left(w_{t+h-i}w_{s+h-j}w_{t-k}w_{s-\ell}\right),$$

and after many messy computations, that yields the asymptotic formula,

$$\mathrm{Var}\left(\tilde{\gamma}_n(h)\right) \approx \frac{1}{n}\left[(\eta-3)\sigma^4|\gamma(h)|^2 + \sum_{\ell=-\infty}^{\infty}\upsilon_\ell\right] \qquad \text{as } n\to\infty.$$

Then one has to appeal to a suitable central limit theorem "for dependent sequences," in order to establish Theorem 19.

Theorem 19 and Slutsky's theorem together yield the following asymptotic normality result for the ACF:

Theorem 21. *Under the conditions of Theorem 19,*

$$\sqrt{n}\left(\hat{\rho}_n(h)-\rho(h)\right) \xrightarrow{d} \mathrm{N}\left(0\,,\tau^2\right) \qquad \textit{as } n\to\infty,$$

where

$$\tau^2 := \sum_{u=1}^{\infty}\left[\rho(u+h)+\rho(u-h)-2\rho(u)\rho(h)\right]^2.$$