Lecture 2: Measures of Dependence and Stationarity

Introduction to Time Series, Fall 2023

Ryan Tibshirani

Related reading: Chapters 1.3–? of Shumway and Stoffer (SS).

1 Mean and variance

- Given a sequence x_t, $t = 1, 2, 3, \dots$, we define its *mean function* (this is viewed as a function of time) by

$$\mu_{x,t} = \mathbb{E}(x_t)$$

  When it is unambiguous from the context which underlying sequence it refers to, we drop the first subscript and simply denote this by μ_t

- Moreover, we define its *variance function* by

$$\sigma^2_{x,t} = \mathrm{Var}(x_t) = \mathbb{E}[(x_t - \mu_t)^2]$$

  Again, when the underlying sequence should be clear from the context, we simplify notation and denote this by σ^2_t

- The mean and variance functions μ_t and σ^2_t are handy objects, because they tell us about salient features of the time series—the drift and spread, respectively, that we should expect over time

- However, in general, they are not enough to characterize the entire distribution of the time series. Why? Two reasons:

  1. In general, the mean and variance are not enough to characterize the *marginal* distribution of a single variate x_t along the sequence
  2. Furthermore, they say nothing about the *joint* distribution of two variates x_s and x_t at different times, $s \neq t$. (For example, do they tend to go up and down together, or do they tend to repel, or ... ?)

  The second of these (joint dependence) we will address soon when we talk about auto-covariance and stationarity. The first (mean and variance specifying the distribution) we will revisit later when we talk about Gaussian processes

- Before moving on though, let's look at some examples. First, let's consider white noise, which recall, refers to a sequence x_t, $t = 1, 2, 3, \dots$ of uncorrelated random variables, with zero mean, and constant variance. Precisely,

$$\mathrm{Cov}(x_s, x_t) = 0, \quad \text{for all } s \neq t$$
$$\mathbb{E}(x_t) = 0, \ \mathrm{Var}(x_t) = \sigma^2, \quad \text{for all } t$$

  So by definition (this one is kind of vacuous), we have mean function $\mu_t = 0$ and variance function $\sigma^2_t = \sigma^2$, which are constant functions (do not vary in time)

- How about a moving average of white noise, with window length 3? This is

$$y_t = \frac{1}{3}\Big(x_{t-1} + x_t + x_{t+1}\Big)$$

Its mean function is

$$\begin{aligned}
\mu_t &= \mathbb{E}(y_t) \\
&= \frac{1}{3}\Big(\mathbb{E}(x_{t-1}) + \mathbb{E}(x_t) + \mathbb{E}(x_{t+1})\Big) \\
&= \frac{1}{3}(0+0+0) \\
&= 0
\end{aligned}$$

Its variance function is

$$\begin{aligned}
\sigma_t^2 &= \mathrm{Var}(y_t) \\
&= \frac{1}{9}\Big(\mathrm{Var}(x_{t-1}) + \mathrm{Var}(x_t) + \mathrm{Var}(x_{t+1}) + \\
&\qquad\quad 2\,\mathrm{Cov}(x_{t-1}, x_t) + 2\,\mathrm{Cov}(x_t, x_{t+1}) + 2\,\mathrm{Cov}(x_{t-1}, x_{t+1})\Big) \\
&= \frac{1}{9}(\sigma^2 + \sigma^2 + \sigma^2 + 0 + 0 + 0) \\
&= \frac{1}{3}\sigma^2
\end{aligned}$$

So its variance is smaller than that of original sequence. In short, smoothing reduces variance

- This last example might have helped you de-rust on some basic facts about expectations and variances. Recall, for constants a_i and random variables x_i:

$$\mathbb{E}\left(\sum_{i=1}^n a_i x_i\right) = \sum_{i=1}^n a_i \mathbb{E}(x_i)$$

$$\mathrm{Var}\left(\sum_{i=1}^n a_i x_i\right) = \sum_{i=1}^n a_i^2\,\mathrm{Var}(x_i) + 2\sum_{i<j}\mathrm{Cov}(x_i, x_j)$$

- The last rule can be thought of as a special case of the more general rule, for constants a_i, b_j, and random variables x_i, y_j:

$$\mathrm{Cov}\left(\sum_{i=1}^n a_i x_i, \sum_{j=1}^m b_j y_j\right) = \sum_{i,j} a_i b_j\,\mathrm{Cov}(x_i, y_j)$$

(To be clear, the sum on the right-hand side above is taken over $i = 1, \ldots, n$ and $j = 1, \ldots, m$)

- Ok, one last example before moving on: let's consider a random walk with drift,

$$x_t = \delta + x_{t-1} + \epsilon_t$$

for a white noise sequence ϵ_t, $t = 1, 2, 3 \ldots$. Recall, we can equivalently write this as (assuming we start at $x_0 = 0$):

$$x_t = \delta t + \sum_{i=1}^t \epsilon_i$$

From this, we can see that the mean function is

$$\mu_t = \delta t + \sum_{i=1}^t \mathbb{E}(\epsilon_i) = \delta t$$

and the variance function is

$$\sigma_t^2 = \sum_{i=1}^t \mathrm{Var}(\epsilon_i) + 2\sum_{i<j}\mathrm{Cov}(\epsilon_i, \epsilon_j) = \sigma^2 t$$

So both the mean and the variance grow over time, proportionally to t. Figure 1 plots example paths over multiple repetitions, for you to get a sense of this

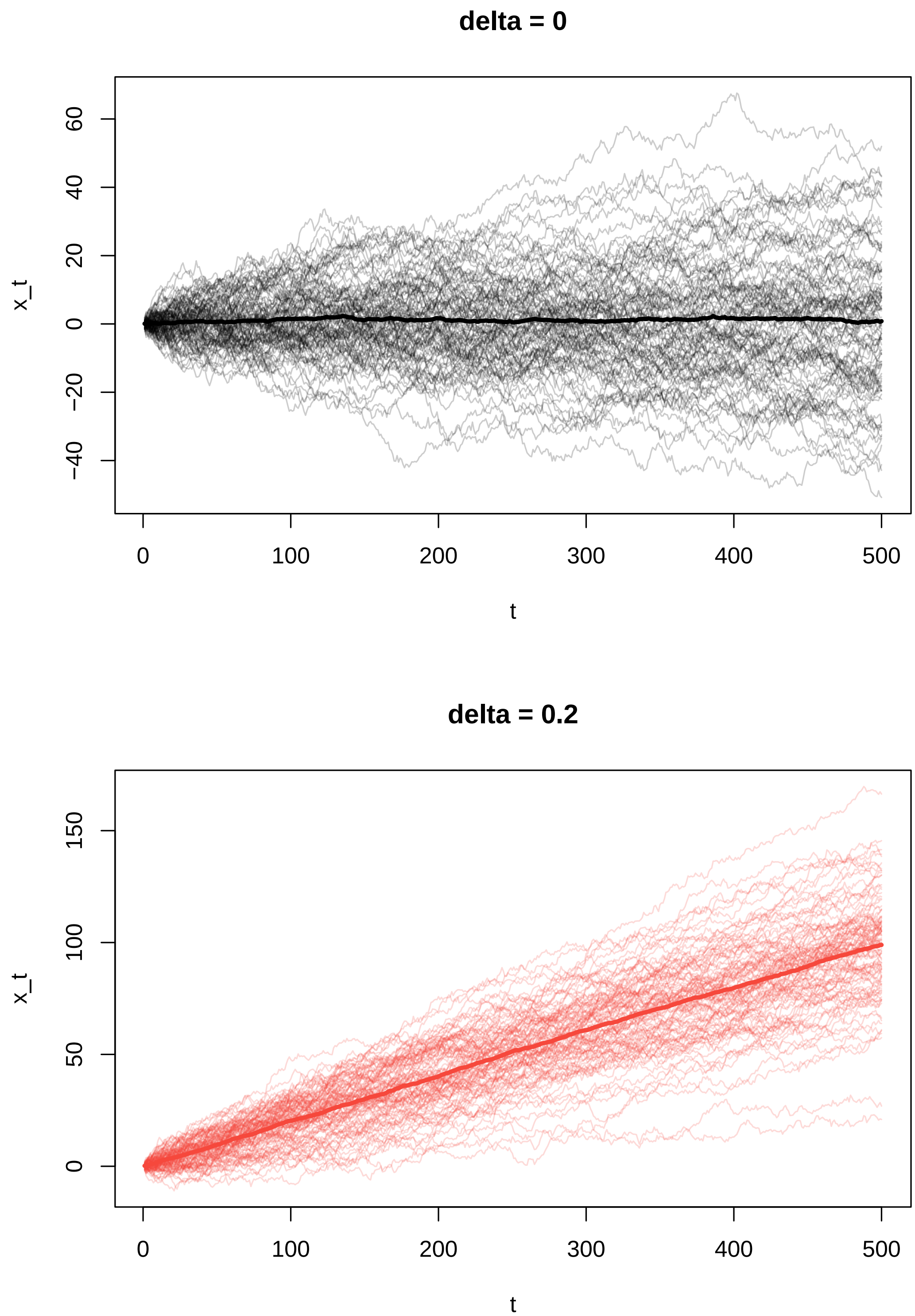

Figure 1: *Random walk without and with drift, each with 100 sample paths. The darker, thicker line in each plot is the sample mean taken at each time point, with respect to the 100 repetitions.*

2 Covariance and correlation

2.1 Auto: one series

- The *auto-covariance function* associated with a time series x_t, $t = 1, 2, 3, \ldots$ is defined as

$$\gamma_x(s,t) = \mathrm{Cov}(x_s, x_t)$$

  This is a symmetric function $\gamma_x(s,t) = \gamma_x(t,s)$, for all s,t. Note that of course $\gamma_x(t,t) = \sigma^2_{x,t}$, the variance function. As before, we will drop the subscript when it is clear from the context what the underlying sequence is, and simply write $\gamma(s,t)$

- The *auto-correlation function* is defined by diving the auto-covariance function by the product of the relevant standard deviations,

$$\rho_x(s,t) = \frac{\gamma_x(s,t)}{\sigma_{x,s}\,\sigma_{x,t}}$$

  which we abbreviate as $\rho(s,t)$ when the underlying sequence is clear from the context

- By the Cauchy-Schwarz inequality, which states that

$$\mathrm{Cov}(x,y) \le \sqrt{\mathrm{Var}(x)\,\mathrm{Var}(y)}$$

  for any random variables x, y, note that we always have

$$\rho_x(s,t) \in [-1, 1]$$

  Typically the auto-correlation will lie strictly in between these limits. (What would a sequence with auto-correlation identically equal to 1 look like? Identically equal to -1?)

- Broadly speaking, the auto-covariance function measures the *linear* dependence between variates along the series. If a series is very smooth, then the auto-covariance function will typically be large (and positive when s,t are close together, but it may be negative when s,t are farther apart). If a series is choppy, then the auto-covariance function will typically be close to zero

- Recall that uncorrelatedness is not the same as independence! So we can have $\gamma_x(s,t) = 0$ for all s,t, even if x_t, $t = 1, 2, 3, \ldots$ are not independent random variables. However, for a Gaussian sequence, uncorrelatedness implies independence

- Let's return to our examples. For white noise, the auto-covariance function is identically zero, $\gamma(s,t) = 0$ for all s,t. Hence the same is true of the auto-correlation function

- For a moving average of white noise, the auto-variance function decreases as the gap between s and t grows. For example, for

$$y_t = \frac{1}{3}\Big(x_{t-1} + x_t + x_{t+1}\Big)$$

  we have

$$\begin{aligned}
\gamma(s,t) &= \mathrm{Cov}(y_s, y_t) \\
&= \mathrm{Cov}\left(\frac{1}{3}\Big(x_{s-1} + x_s + x_{s+1}\Big), \frac{1}{3}\Big(x_{t-1} + x_t + x_{t+1}\Big)\right) \\
&= \begin{cases} \sigma^2/9 & s = t-2 \\ 2\sigma^2/9 & s = t-1 \\ \sigma^2/3 & s = t \\ 2\sigma^2/9 & s = t+1 \\ \sigma^2/9 & s = t+2 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

(You can go through each case carefully, and use the formula for the covariance of linear combinations given previously.) The auto-correlation function simply divides this by σ^2. since the variance function is constant, and is hence

$$\rho(s,t) = \begin{cases} 1/9 & s = t-2 \\ 2/9 & s = t-1 \\ 1/3 & s = t \\ 2/9 & s = t+1 \\ 1/9 & s = t+2 \\ 0 & \text{otherwise} \end{cases}$$

- For a random walk (with or without drift), the auto-covariance function also decreases as the gap between s and t grows, but has a different structure. Considering

$$x_t = \delta t + \sum_{i=1}^{t} \epsilon_i$$

we have

$$\begin{aligned} \gamma(s,t) &= \mathrm{Cov}(x_s, x_t) \\ &= \mathrm{Cov}\left(\delta s + \sum_{i=1}^{s} \epsilon_i, \delta t + \sum_{i=1}^{t} \epsilon_i\right) \\ &= \sigma^2 \min\{s,t\} \end{aligned}$$

(To see this more clearly, consider the case where $s < t$ and recognize that the sums above overlap with exactly s white noise variates.) The auto-correlation function divides this by the product of the relevant variances:

$$\begin{aligned} \rho(s,t) &= \frac{\sigma^2 \min\{s,t\}}{\sigma\sqrt{s} \cdot \sigma\sqrt{t}} \\ &= \frac{\min\{s,t\}}{\sqrt{st}} \end{aligned}$$

- Figure 2 gives a visualization of the auto-correlation functions for the moving average and random walk settings. The moving average auto-correlation function is presented as a banded matrix (though it is hard to see the band since the sequence is of total length $n = 500$ and most values in the auto-correlation matrix are zero). Importantly, we can see that the same pattern persists throughout the whole matrix, and all that matters is the distance to the diagonal. This is an important property that we will revisit soon (hint: stationarity). Meanwhile, the random walk auto-correlation function does *not* have a pattern than persists throughout, and we can see a “cone” that grows around the diagonal as time grows

2.2 Cross: two series

- The *cross-covariance function* associated with two time series x_t, $t = 1,2,3,\dots$ and y_t, $t = 1,2,3,\dots$ is defined as

$$\gamma_{xy}(s,t) = \mathrm{Cov}(x_s, y_t)$$

This is *not* necessarily a symmetric function, and generically $\gamma_{xy}(s,t) \neq \gamma_{xy}(t,s)$. Note that the cross-covariance between a time series and itself is simply its auto-covariance, i.e., $\gamma_{xx}(t,t) = \gamma_x(t)$

- The *cross-correlation function* is defined by diving the cross-covariance function by the product of the relevant standard deviations,

$$\rho_{xy}(s,t) = \frac{\gamma_{xy}(s,t)}{\sigma_{x,s}\, \sigma_{y,t}}$$

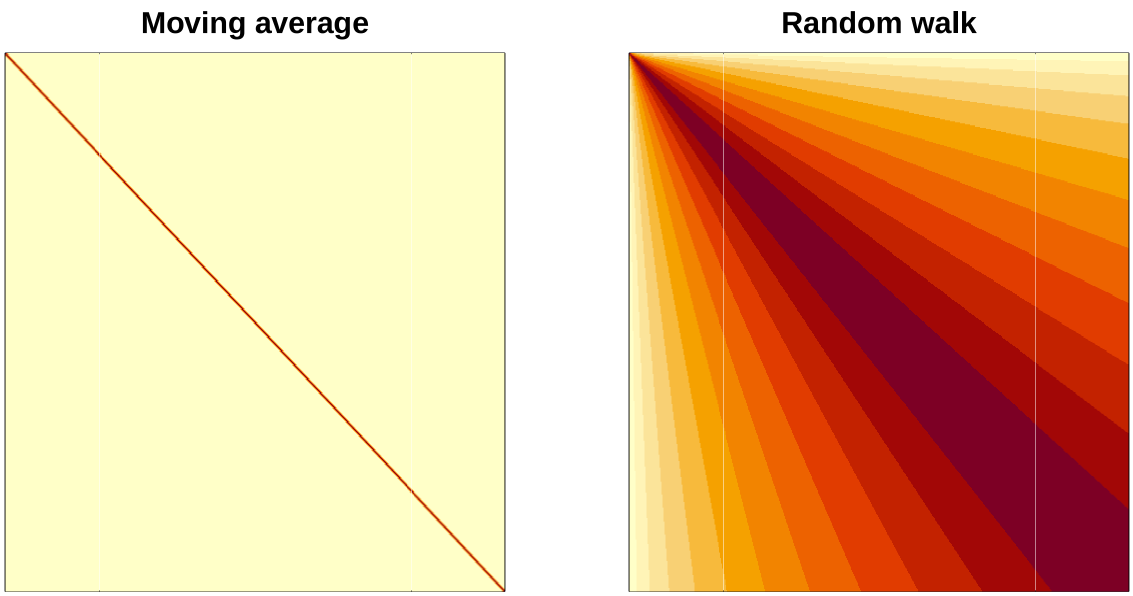

Figure 2: *Heatmaps of the auto-correlation functions for the moving average and random walk examples, for a sequence with* $n = 500$ *time points. The heatmaps are laid out in the same way that we would naturally view a matrix:* $(s, t) = (1, 1)$ *is the top left corner, and* $(s, t) = (n, n)$ *is the bottom right (that is,* s *increases along the rows, and* t *increases along the columns). Yellow reflects a value of zero, and darker red reflects a larger value.*

By Cauchy-Schwarz, once again, we know that $\rho_{xy}(s, t) \in [-1, 1]$

- Figure 4 shows an example of an *estimated* cross-correlation function for Covid-19 cases (the first series x_s) and deaths (the second series y_t) in California, which are plotted in Figure 3. We can see that the cross-correlation is plotted as a function of "lag", which refers to the value $h = s - t$, and appears to be maximized at a lag of $h = -25$ or so. This makes sense, in that we expect cases to be highly correlated with deaths several weeks later (this is also visually apparent in Figure 3)

  But wait a minute ... why have we reduced the whole cross-correlation function, which is generically a function of two time indexes s and t, to be a function of a single number, lag, $h = t - s$? Because that is really the only way it is estimable (unless we have more information than the two time series at hand). More on this shortly, but next, we'll cover stationarity, which will provide the foundation for this estimation strategy in the first place

3 Stationarity

3.1 Strong

3.2 Weak

4 Estimation

5 Gaussian processes

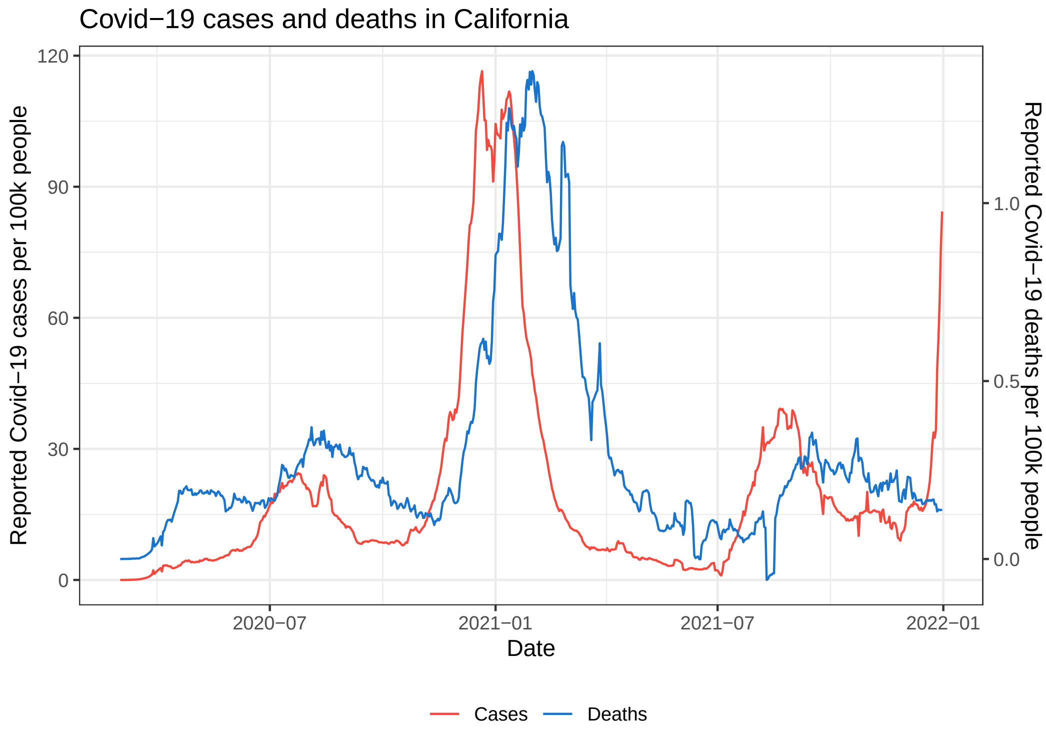

Figure 3: *Covid-19 cases and deaths, in the state of California.*

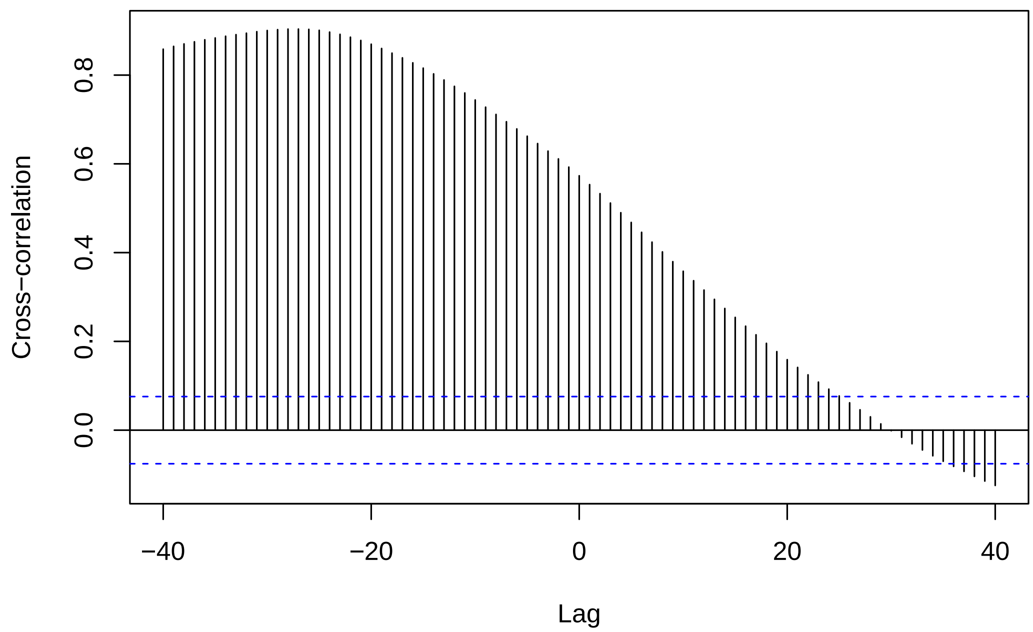

Figure 4: *Cross-correlation function for Covid-19 cases and deaths in California, as plotted above. This is estimated by the* `ccf()` *function in R.*