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, $t = 1, 2, 3, \ldots$, 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
$$\sigma_{x,t}^2 = \text{Var}(x_t) = \mathbb{E}[(x_t - \mu_t)^2]$$
Again, when the underlying sequence should be clear from the context, we simplify notation and denote this by σ_t^2

- The mean and variance functions μ_t and σ_t^2 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 autocovariance 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, $t = 1, 2, 3, \ldots$ of uncorrelated random variables, with zero mean, and constant variance. Precisely,
$$\text{Cov}(x_s, x_t) = 0, \quad \text{for all } s \neq t$$
$$\mathbb{E}(x_t) = 0, \ \text{Var}(x_t) = \sigma^2, \quad \text{for all } t$$
So by definition (this one is kind of vacuous), we have mean function $\mu_t = 0$ and variance function $\sigma_t^2 = \sigma^2$, 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

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

Its variance function is

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

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_t:

$$\mathbb{E}\left(\sum_{t=1}^{k} a_t x_t \right) = \sum_{t=1}^{k} a_t \mathbb{E}(x_t)$$

and

$$\mathrm{Var}\left(\sum_{t=1}^{k} a_t x_t \right) = \sum_{s,t} a_s a_t \,\mathrm{Cov}(x_s, x_t)$$
$$= \sum_{t=1}^{k} a_t^2 \,\mathrm{Var}(x_t) + 2 \sum_{s<t} \mathrm{Cov}(x_s, x_t)$$

The last rule can be thought of as a special case of the more general rule, for constants a_t and b_t:

$$\mathrm{Cov}\left(\sum_{t=1}^{k} a_t x_t, \sum_{t=1}^{\ell} b_t y_t \right) = \sum_{s,t} a_s b_t \,\mathrm{Cov}(x_s, y_t)$$

(To be clear, the sum on the right-hand side above is taken over $s = 1, \ldots, t$ and $t = 1, \ldots, \ell$)

- 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

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

2

and the variance function is

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

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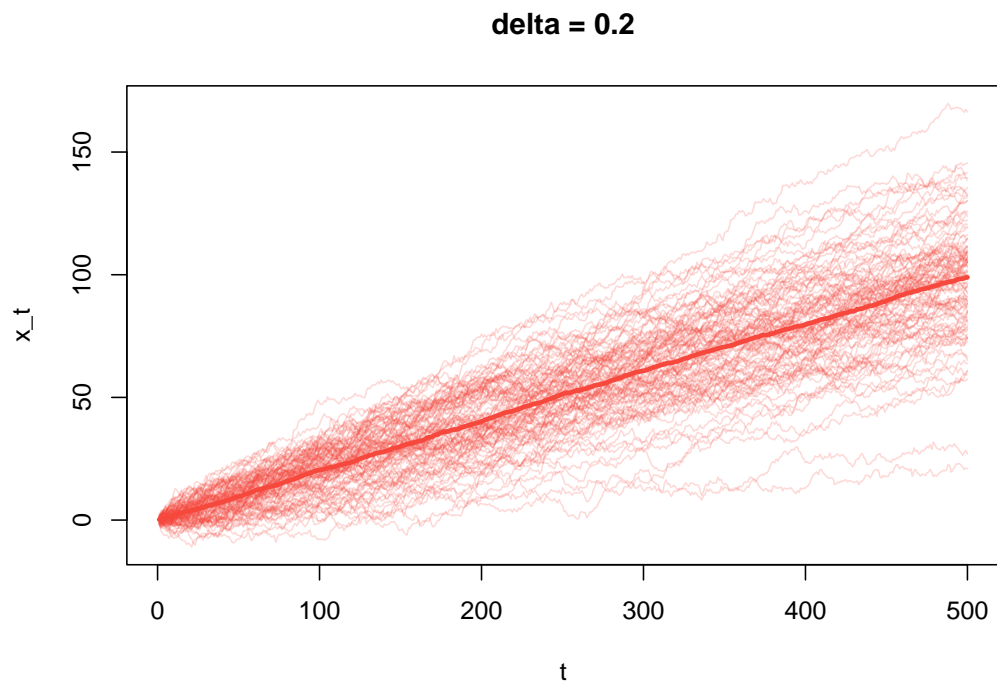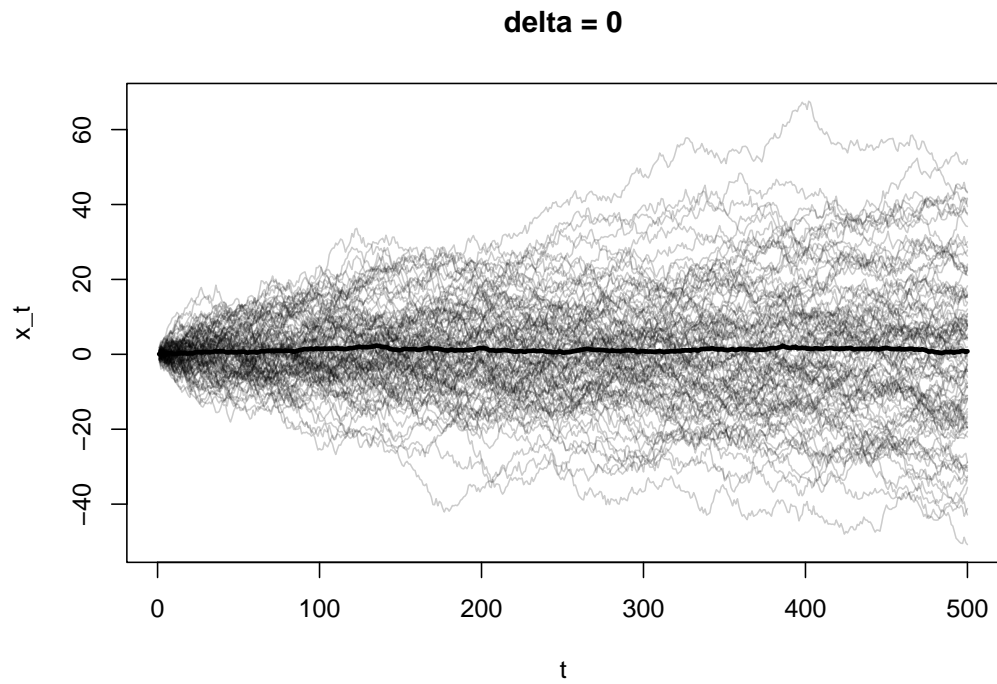
Figure 1: *Random walk without and with drift, each with 100 sample paths. The darker, thicker line in each plot is the sample mean taken at with time point, with respect to the 100 repetitions.*