Contents

(All variables are real and one-dimensional unless otherwise specified.)

1. Autocorrelation

Let $X = \{X_t \mid t \geq 1\}$ be a time series. Its **autocorrelation** with a **lag** τ is defined as

$$R_X(\tau) = \langle X_t X_{t+\tau} \rangle \,.$$

Intuitively it indicates the correlation of data separated τ units apart. If X_t and $X_{t+\tau}$ are often both high or low, $R_X(\tau)$ is high. ("Auto-correlation" literally means the correlation with one's self.) Since "high" and "low" are relative to the mean μ_X of the series, one may redefine autocorrelation with

$$K_X(\tau) = \langle (X_t - \mu_X)\,(X_{t+\tau} - \mu_X) \rangle = R_X(\tau) - \mu_X^2 \,.$$

Here comes a trouble in naming: while physicists prefer still calling it "autocorrrelation", statisticians prefer calling this new quantity "**autocovariance**" instead. Worse still, the **autocorrelation coefficient** of X is indisputably always defined as the normalized version of $K_X(\tau)$, i.e.

$$\rho_X(\tau) = \frac{K_X(\tau)}{K_X(0)} = \frac{K_X(\tau)}{\sigma_X^2} \,,$$

where $K_X(0) \equiv \sigma_X^2$ is the variance of X. Because $\sigma_X^2 \equiv K_X(0)$, it also makes sense to regard as

1.1 Discrete and continuous time series

The "time" of a time series may be either **discrete** ($t \in \mathbb{Z}$) or **continuous** ($t \in \mathbb{R}$). On one hand, if time is discrete, a time average normalizes a **sum** over time, so $\langle X_t X_{t+\tau} \rangle \sim \sum_t X_t X_{t+\tau}$. On the other hand, if time is continuous, a time average normalizes an **integral** over time instead, so $\langle X_t X_{t+\tau} \rangle \sim \int X_t X_{t+\tau} \mathrm{d}t$.

Although the time that we can practically measure is always discrete, we may want to approximate it with a continuous variable to simplify some algebra.

1.2 Finitely and infinitely long time series

The mean and the variance of a time series are, strictly speaking, meaningless unless the time series is **infinitely long**, which we can never obtain in reality. Instead, for a T-unit **finitely long** time series, we have to estimate the two entities with their sample counterparts, viz. the **sample mean** $\hat{\mu}_X = \frac{1}{T} \sum_t X_t$ and the **sample variance** $\hat{\sigma}_X^2 = \frac{1}{T-1} \sum_t (X_t - \hat{\mu})^2$.

As time is now bounded in $[1, T]$, we must ensure that $t + \tau \leq T$, resulting in $\langle X_t X_{t+\tau} \rangle \sim \sum_{t=1}^{T-\tau} X_t X_{t+\tau}$. **Here comes a tricky point: what should the normalization constant be?** Although there are $T - \tau$ terms in the summation, it is **preferred** (i.e. not mandated) to normalize the sum with $\frac{1}{T}$ instead of $\frac{1}{T-\tau}$. The justification behind this is quite involved. A simplified explanation is that for a long-term correlation with $\tau \to T$, the summation only sums up a few terms and thus becomes highly prone to noise, which should be suppressed by a harder normalization constant . As long as a short-term correlation with $\tau \ll T$ is concerned, $\frac{1}{T}$ barely differs from $\frac{1}{T-\tau}$.

1.3 Example: a cosine signal

Consider a continuous-time cosine signal $X(t) = A\cos(\omega t + \phi) + c$ with $t \in [0, \infty)$, where A, ω, ϕ, and c are arbitrary constants. What are its mean, variance, and autocorrelation?

Solution. The mean is

$$\begin{aligned} \mu_X &= c + \lim_{T\to\infty} \frac{A}{T} \int_0^T \cos(\omega t + \phi) \mathrm{d}t \\ &= c + \lim_{T\to\infty} \frac{A}{\omega T} [\sin(\omega T + \phi) - \sin\phi] \ . \end{aligned}$$

As $\sin(\omega T + \phi) - \sin\phi$ is bounded in $[-2, 2]$, the limit is squeezed by $\frac{1}{T}$ to be 0, leaving $\mu_X = c$.

Then, the autocorrelation is

$$R_X(\tau) = \langle A^2 \cos(\omega t + \phi) \cos(\omega t + \omega\tau + \phi) + Ac\cos(\omega t + \phi) + Ac\cos(\omega t + \omega\tau + \phi) + c^2 \rangle \,,$$

where the second and the third terms also vanish due to squeezing. By the product-to-sum formula, the first term becomes

$$\lim_{T\to 0} \left[\frac{A^2}{2T} \int_0^T \cos(\omega\tau)\mathrm{d}t + \frac{A^2}{2T} \int_0^T \cos(2\omega t + \omega\tau + 2\phi)\mathrm{d}t \right] = \frac{A^2}{2}\cos(\omega\tau) \,,$$

where the second integral vanishes, again, due to squeezing, so

$$R_X(\tau) = \frac{A^2}{2}\cos(\omega\tau) + c^2 \,.$$

Finally, the variance is simply $\sigma_X^2 = R_X(0) = \frac{A^2}{2} + c^2$.

2. Moving average

It is often assumed that a time series $X_t = \tilde{X}_t + \varepsilon_t$ consists of an underlying pattern $\tilde{X}$ and some random noise ε that satisfies $\langle \varepsilon \rangle \to 0$. While noise distorts a time series, **moving average** helps attenuate the noise and thus reveal its underlying pattern. There are various kinds of moving average. I am going to introduce three.

2.1 Simple moving average

The simple moving average (SMA) of X reads

$$\mathrm{SMA}_X(t; w) = \frac{1}{w} \sum_{i=0}^{w-1} X_{t-i} \,,$$

where the parameter w corresponds to a **window** size. SMA averages out every data point with its $w-1$ previous data points. (The points are thus said to be in the same window.) In this way, SMA helps smooth out noise with a period shorter than w. For example, a sinusoidal noise $\varepsilon_t = \sin t$ disappears in a window with $w = 2\pi$. However, the underlying pattern is also blurred if the wind is too long.

2.2 General moving average

Imagine a stock: everyone expects that its price today receives a stronger influence from yesterday's headlines than a financial crisis decades ago. It then makes sense to define a general moving average (GMA) that **emphasizes new data more than old data**.

There are various ways to execute this idea. The lecture suggests

$$\mathrm{GMA}_X(t; w, \tau) = \frac{1}{Z_{w,\tau}} \sum_{i=0}^{w-1} X_{t-i} e^{-i/\tau} ,$$

which has a window size w, a **memory** length $\tau \in (0, \infty)$, and a normalization constant $Z_{w,\tau} = \sum_{i=0}^{w-1} e^{-i/\tau}$. When τ increases, the old data becomes as important as the new data; when $\tau \to \infty$, the definition is reduced to the SMA.

2.3 Exponential moving average

The exponential moving average (EMA) also emphasizes new data more than old data. It has a recursive definition:

$$\begin{aligned}\mathrm{EMA}_X(t;a) =&(1-a)\,\mathrm{EMA}_X(t-1;a)+aX_{t-1}\\ =&a\sum_{i=0}^{t-1}X_{t-i}(1-a)^i\,,\end{aligned}$$

where $a\in[0,1]$ is some attenuation factor. A larger a puts a greater emphasis on new data. To simulate the effect of a w-unit long window with EMA, an intuitive choice is $a=\frac{1}{w}$, which suppresses a data point's weight by $\frac{1}{e}$ every w units, but a more common choice turns out to be $a=\frac{2}{w+1}$.

Despite their identical spirit, EMA is much more prevalent than GMA because of it simpler definition and more efficient computation: EMA only costs $O(1)$ once its previous value is given, whereas GMA always costs $O(w)$.

Still, techinally speaking, only GMA is a moving average, while EMA should be classified as **autoregression**. This dichotomy arises because they treat data in fundamentally distinct ways. Suppose X jumps sharply at $t=1$ due to an impulse $\varepsilon_t=\delta(t-1)$. The impulse no longer influences GMA once it leaves the window, i.e. $t>1+w$; in contrast, the remnant of the impulse, however small it is, stays in EMA forever. (Therefore, engineers alternatively call GMA a "**finite-impulse-response** filter" and EMA an "**infinite-impulse-response** filter".)

3. Technical analysis

In trading, the way of predicting a stock's price solely based on its previous prices is called **technical analysis**, for which MACD, standing for **moving-average-convergence-divergence**, is an important tool. If X represents the time series of a stock's daily price, its **MACD line** (aka MACD proper) is defined as

$$M_X(t) = \mathrm{EMA}_X(t; w = 12) - \mathrm{EMA}_X(t; w = 26)\,.$$

Then its corresponding **signal line** is given as

$$S_X(t) = \mathrm{EMA}_{M_X}(t; w = 9)\,.$$

Remember that EMA is defined with an attenuation factor a instead of a window size w, so you need to decide how to express a in terms of w. Meanwhile, the three values of w (viz. 9 days, 12 days, and 26 days) are purely conventional: they are chosen just because if a stock market has six working days per week, the periods will respectively correspond to 1.5 weeks, 2 weeks, and 1 month.

Technical analysis contrasts with **fundamental analysis**, which aims at predicting a stock's price based on other indicators of its company's perfomance like financial statements. Whether technical analysis is valid, scientific, or possible at all remains an open question.

3.1 Interpretation

Traders usually interprets the MACD of a stock's price X from three aspects. In their jargon, "bullish" and "bearish" means an increasing trend and a decreasing trend respectively as a bull's head points upwards but a bear's head points downwards.

- **Signal-line crossover.**
  - When M_X breaks above S_X, there is a **strong bullish sign**, suggesting that X may **rise rapidly**.
  - In contrast, when the former breaks below the latter, there is a **strong bearish sign**, suggesting that X may **drop rapidly**.
- **Zero crossover.**

- When M_X breaks above zero, there is a **mild bullish sign**, suggesting that X may **rise moderately**. This is also called a **golden cross**.
- In contrast, when it breaks below zero, there is a **mild bearish sign**, suggesting that X may **drop moderately**. This is also called a **death cross**.

- **Divergence.**
  - When X drops to a new low but M_X does not, there is a **sign of bullish divergence**, suggesting that X may stop dropping.
  - In contrast, when the former rises to a new high but the latter does not, there is a **sign of bearish divergence**, suggesting that X may stop rising.

3.2 Why does MACD work?

Traders uses MACD because it works empirically. But why does it work? Loosely speaking, if X represents a random walker's **displacement**, M_X and $M_X - S_X$ can be respectively treated as his **velocity** and **acceleration**. The ideas introduced above become more sensible once you use a perspective of physics.

Yet, why are the financial things comparable to physics? The mathematical explanation turns out to be quite difficult in my opinion. If you are interested, please visit this webpage.