Seminar 2: Correlation and WLLN

Carlos Buitrago

Table of contents

In many applications we are interested not only in the behavior of a single random variable, but also in the relationship between two random variables.

Covariance measures the direction of the linear relationship between two variables, while the correlation coefficient gives a normalized measure that always lies between -1 and 1.

These quantities are fundamental throughout statistics, machine learning, finance, econometrics, and data analysis.

1 Covariance

Let ξ and η be two random variables with finite expectations.

The **covariance** between ξ and η is defined as

$$\mathrm{Cov}(\xi, \eta) = \mathbb{E}[(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)] .$$

Using linearity of expectation,

$$\boxed{\mathrm{Cov}(\xi, \eta) = \mathbb{E}[\xi\eta] - \mathbb{E}[\xi]\,\mathbb{E}[\eta].}$$

Intuitively,

- positive covariance means that both variables tend to increase together;
- negative covariance means that one variable tends to increase while the other decreases;
- covariance close to zero indicates the absence of a linear relationship (although nonlinear dependence may still exist).

1.1 Properties of Covariance

For any random variables ξ, η, ξ_1, ξ_2 and constants a, b,

1. **Symmetry**

$$\mathrm{Cov}(\xi, \eta) = \mathrm{Cov}(\eta, \xi).$$

2. **Covariance with itself**

$$\mathrm{Cov}(\xi, \xi) = \mathrm{Var}(\xi).$$

3. **Linearity**

$$\mathrm{Cov}(a\xi_1 + b\xi_2, \eta) = a\ \mathrm{Cov}(\xi_1, \eta) + b\ \mathrm{Cov}(\xi_2, \eta).$$

4. **Independence implies zero covariance**

   If $\xi \perp\!\!\!\perp \eta$, then

$$\mathrm{Cov}(\xi, \eta) = 0.$$

   However, the converse is generally false: zero covariance does **not** imply independence.

1.2 Variance of a Sum

One of the most useful applications of covariance is computing variances of sums.

For any random variables $\xi_1, \dots, \xi_n$,

$$\boxed{\mathrm{Var}\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(\xi_i, \xi_j).}$$

Equivalently,

$$\mathrm{Var}\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \mathrm{Var}(\xi_i) + 2 \sum_{i<j} \mathrm{Cov}(\xi_i, \xi_j).$$

If the variables are independent, all covariance terms vanish, giving

$$\mathrm{Var}\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \mathrm{Var}(\xi_i).$$

2 Correlation Coefficient

Since covariance depends on the units of measurement, it is difficult to compare across different datasets.

The **correlation coefficient** normalizes covariance by the standard deviations:

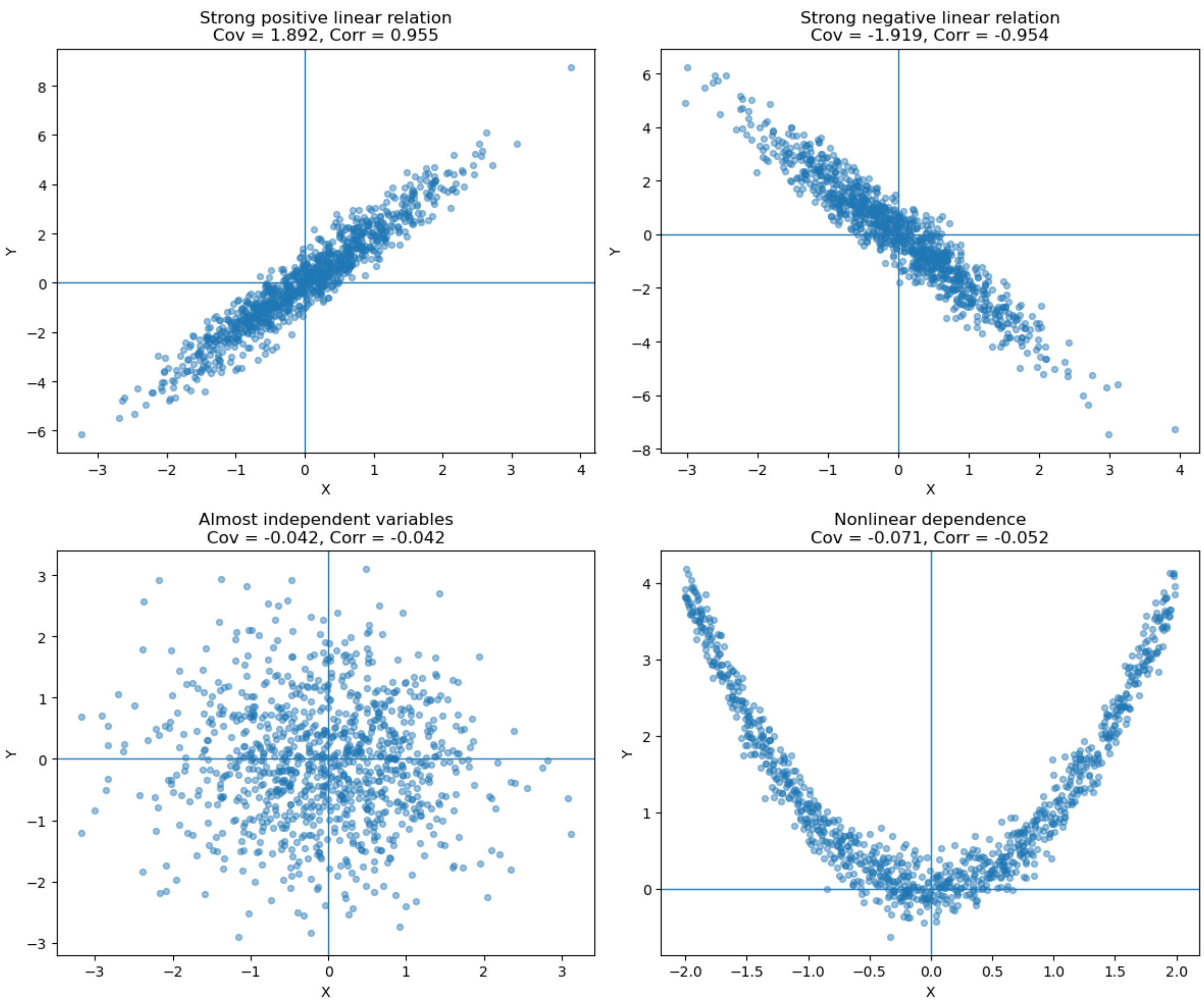
$$\boxed{\mathrm{Corr}(\xi,\eta)=\frac{\mathrm{Cov}(\xi,\eta)}{\sqrt{\mathrm{Var}(\xi)}\sqrt{\mathrm{Var}(\eta)}}.}$$

The correlation coefficient is dimensionless and satisfies

$$-1\le \mathrm{Corr}(\xi,\eta)\le 1.$$

Its interpretation is straightforward:

- $\mathrm{Corr}=1$: perfect positive linear relationship;
- $\mathrm{Corr}=-1$: perfect negative linear relationship;
- $\mathrm{Corr}=0$: no linear relationship.



2.1 Example: GDP Growth and Association Between Two Countries

The following data (this is not real data BTW) represent GDP growth indicators (in percent relative to the year 2005) for **Russia** and **Belarus** over the same 14-year period.

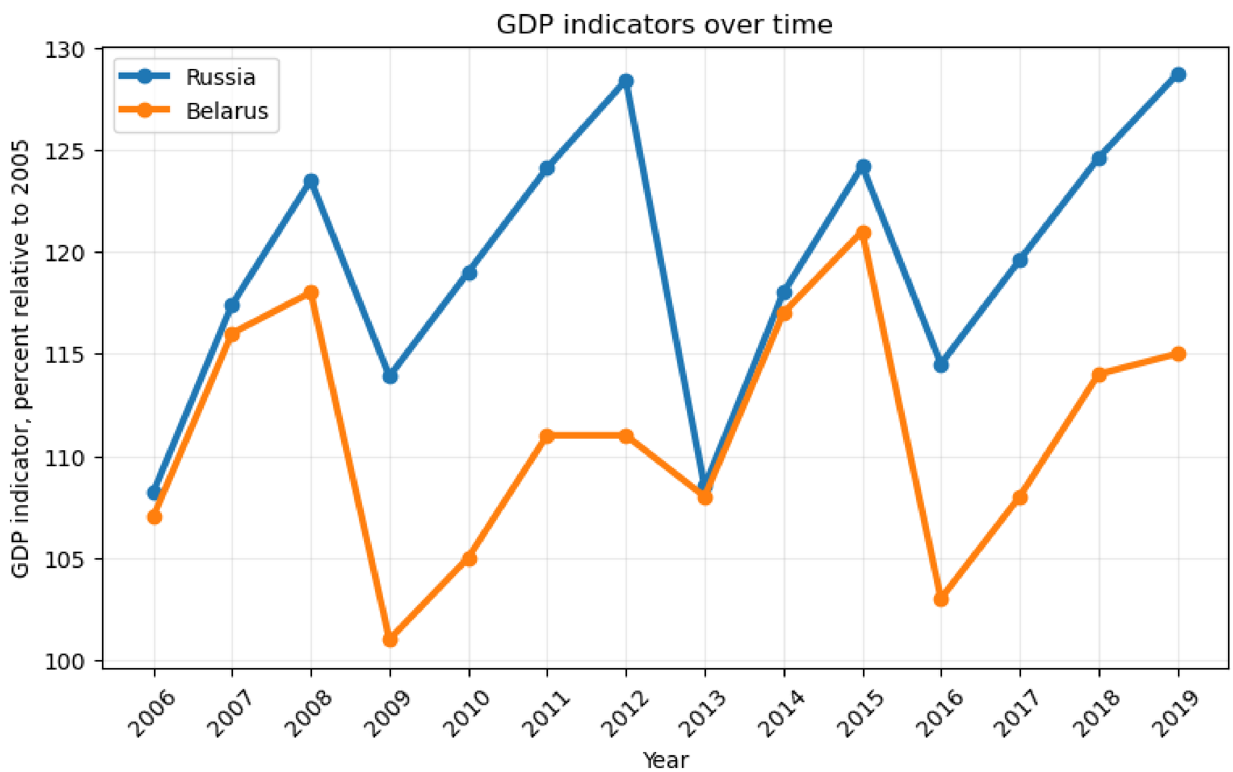
Russia

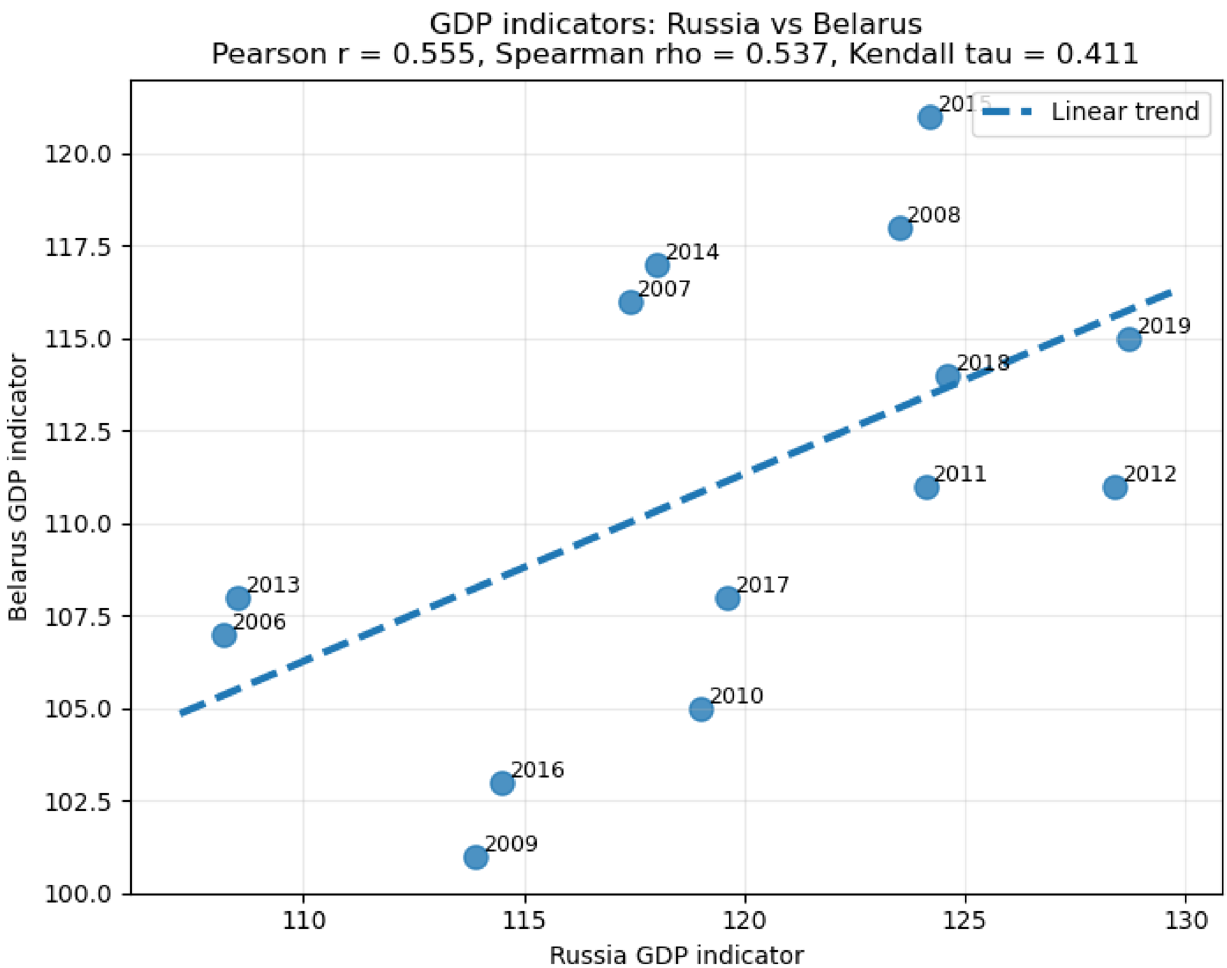
$$108.2,\ 117.4,\ 123.5,\ 113.9,\ 119.0,\ 124.1,\ 128.4,\ 108.5,\ 118.0,\ 124.2,\ 114.5,\ 119.6,\ 124.6,\ 128.7$$

Belarus

$$107,\ 116,\ 118,\ 101,\ 105,\ 111,\ 111,\ 108,\ 117,\ 121,\ 103,\ 108,\ 114,\ 115$$

Our goal is to determine whether these two samples appear to be statistically independent.





3 Concentration Inequalities

In many situations the exact distribution of a random variable is unknown or difficult to compute.

Instead of finding exact probabilities, we seek **upper bounds** for probabilities of large deviations from typical values.

3.1 Markov's Inequality

Let $\xi \geq 0$ be a nonnegative random variable with finite expectation.

Then for every $a > 0$,

$$\boxed{\mathbb{P}(\xi \geq a) \leq \frac{\mathbb{E}[\xi]}{a}.}$$

This inequality requires only the existence of the expectation and makes no assumptions on the distribution. Although often crude, it serves as the starting point for many stronger inequalities.

3.2 Chebyshev's Inequality

Suppose ξ has finite expectation and finite variance.

Then for every $\varepsilon > 0$,

$$\boxed{\mathbb{P}\left(|\xi - \mathbb{E}\xi| \geq \varepsilon\right) \leq \frac{\text{Var}(\xi)}{\varepsilon^2}.}$$

Unlike Markov's inequality, Chebyshev's inequality uses both the expectation and the variance.

4 Weak Law of Large Numbers

Suppose $X_1, X_2, \ldots$ are independent and identically distributed random variables satisfying

$$\mathbb{E}[X_i] = \mu, \qquad \text{Var}(X_i) = \sigma^2 < \infty.$$

The sample mean is $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

The Weak Law of Large Numbers states that

$$\boxed{\overline{X}_n \xrightarrow{P} \mu.}$$

In other words,

$$\lim_{n\to\infty} \mathbb{P}\left(|\overline{X}_n - \mu| > \varepsilon\right) = 0 \qquad \text{for every } \varepsilon > 0.$$

4.1 Bernoulli Trials and the Interpretation of Probability

One of the most important consequences of the Weak Law of Large Numbers is that it provides a mathematical justification for interpreting **probability as long-run relative frequency**.

Suppose we repeat the same experiment independently many times. Let

$$X_i = \begin{cases} 1, & \text{if the event occurs on trial } i, \\ 0, & \text{otherwise.} \end{cases}$$

Then $X_i \sim \text{Bernoulli}(p)$. Therefore

$$\mathbb{E}[X_i] = p, \qquad \text{Var}(X_i) = p(1-p).$$

Notice that $\sum_{i=1}^n X_i$ is exactly the number of times the event occurred during the first n trials. Therefore,

$$\overline{X}_n = \frac{\text{Number of successes}}{\text{Number of trials}},$$

which is precisely the **empirical frequency** (or relative frequency) of the event.

Applying the Weak Law of Large Numbers,

$$\overline{X}_n \xrightarrow{P} p,$$

or equivalently,

$$\boxed{\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\text{Number of successes}}{n} - p\right| > \varepsilon\right) = 0 \qquad (\varepsilon > 0).}$$

Thus, as the number of independent repetitions increases, the observed frequency of the event becomes arbitrarily close to its true probability with probability approaching one.

This result gives a rigorous mathematical foundation for the frequentist interpretation of probability:

> **The probability of an event is the limiting value of its relative frequency in a large number of independent repetitions of the same experiment.**