Topic: Intro to Stochastic Processes

Prof. Victor M. Preciado

Contents

1 Time-Series Data

Time-series data is prevalent in many fields, including:

- ***Finance:*** Stock prices, exchange rates, and economic indicators.
- ***Signal Processing:*** Audio signals, radar signals, and communication signals.
- ***Engineering:*** Sensor data in manufacturing, energy consumption patterns, etc.
- ***Healthcare:*** Patient vital signs, such as heart rate and blood pressure, monitored over time.
- ***Weather Forecasting:*** Temperature, precipitation, and other meteorological variables.

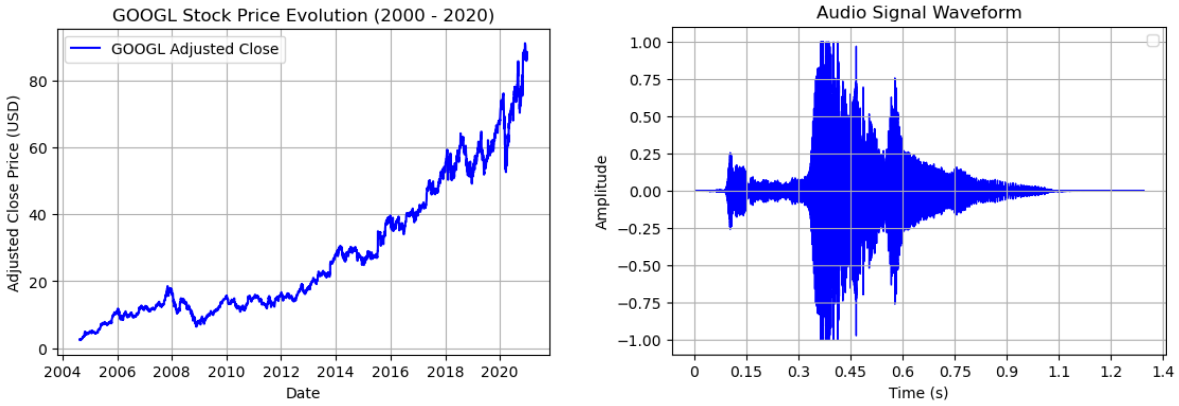

Figure 1: (Left) Evolution of Google Stock prices from 2000 to 2020. (Right) Waveform of a sneeze.

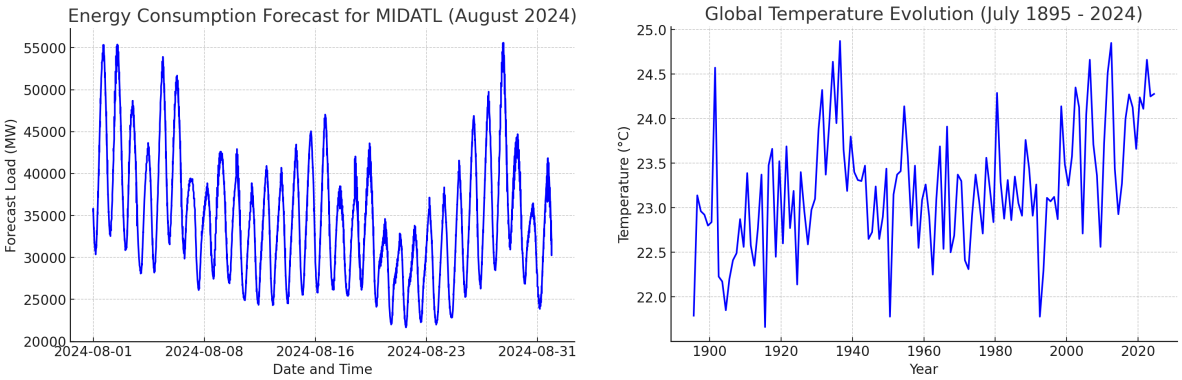

Figure 2: (Left) Energy consumption in the Mid-Atlantic U.S. region during August 2024. (Right) Global temperature evolution from 1895 to 2024

Given the widespread presence of time series data across these diverse domains, our ability to analyze and interpret such data is of paramount importance. Tools for time series analysis allow practitioners to *uncover patterns, identify trends, and make forecasts*, which are critical for informed decision-making. It is worth noting that certain phenomena are inherently more predictable than others. Our ability to forecast a particular event or value depends on several critical factors, such as

the quantity and quality of the available data. A central aspect of effective forecasting is determining when accurate predictions are feasible, as opposed to when forecasts offer no advantage over random chance. Reliable forecasts capture authentic patterns and relationships within historical data, without simply replicating past events that are unlikely to recur. In this context, distinguishing between random fluctuations that should be disregarded and genuine patterns that require modeling is key. In this direction, the complexity of time series data requires specialized techniques to extract meaningful insights.

1.1 The Forecasting Process

The forecasting process typically involves the following steps:

1. **Gathering Information:** Two types of information are essential: *statistical data* and *expert knowledge* from those familiar with the system. Expert knowledge helps identify the underlying causal factors, while the data collected must be statistically informative, meaning past observations should provide insights into future outcomes.

2. **Preliminary Analysis:** Graphing the data helps identify trends, seasonality, and potential outliers. It also provides insight into relationships between variables and helps guide model selection.

3. **Model Selection and Fitting:** The choice of model depends on the availability of data, the strength of relationships between variables, and the specific objectives of the forecast. In this course, we will explore a diverse array of models, each with its own underlying assumptions and methodological approaches. Understanding these assumptions is crucial, as they guide the model's applicability to different types of data and forecasting scenarios.

4. **Model Evaluation:** Once a model has been selected, forecasts are generated and their accuracy is assessed as actual data becomes available. In this course, we will cover several techniques to evaluate forecast accuracy, while practical challenges such as missing data and limited time series length must be carefully managed during implementation.

The rigorous analysis and forecasting of time series data require a deep understanding of the underlying stochastic nature of the data. This involves treating time series as random processes, where each observation is seen as a realization of a stochastic process. By exploring time series through this lens, we can better model, predict, and infer patterns that are otherwise obscured by the inherent randomness in the data.

2 Time Series as Stochastic Processes

A time series of length L is a sequence of observations $(y_1, y_2, \ldots, y_L)$ recorded at specific time points $t_1, t_2, \ldots, t_L$. In this course, we will focus on the case in which these time points are uniformly separated, i.e., $t_k - t_{k-1} = T$ for all $k \in \mathbb{N}$, with T being a constant **sampling period**. Statistically, the entries of a time series can be interpreted as realizations of a sequence of random variables[1] $(Y_1, Y_2, Y_3, \ldots)$. This means that the value y_k observed at time t_k can be interpreted as a realization of the random variable Y_k. The entire collection of these random variables, indexed by time, forms

[1] All these random variables are defined on the same probability space $(\Omega, \mathcal{F}, P)$.

what is known as a discrete-time **stochastic process**, denoted by $\mathcal{Y} = \{Y_k : k \in \mathbb{N}\}$. The observed sequence $(y_1, y_2, \ldots, y_L)$ is referred to as a **sample path** of the underlying stochastic process. This statistical perspective enables the use of probabilistic methods to analyze and forecast the behavior of time series data.

2.1 Statistical Properties of Stochastic Processes

Understanding the statistical properties of a random process is crucial in time series analysis, as it allows for rigorous modeling, prediction, and inference by capturing the underlying patterns and dependencies in the data. A complete probabilistic description of a stochastic process $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_L\}$ of length L is provided by its **Joint Distribution Function (JDF)**. The JDF is a multivariate CDF defined as:

$$F_{\mathcal{Y}}(y_1, \ldots, y_L) = \Pr\{Y_1 \leq y_1, \ldots, Y_L \leq y_L\}.$$

Although the joint distribution function (JDF) fully describes the stochastic process, it is often difficult to work with and not easily accessible for time series analysis. From the JDF, one can extract the marginal distributions of each random variable in the stochastic process, denoted by $F_{Y_k}(y_k) = \Pr\{Y_k \leq y_k\}$, as well as the joint pairwise distribution for each pair, denoted by $F_{Y_{k_1}, Y_{k_2}}(y_{k_1}, y_{k_2}) = \Pr\{Y_{k_1} \leq y_{k_1}, Y_{k_2} \leq y_{k_2}\}$.

The mean, variance, autocovariance, and autocorrelation are fundamental statistical properties of a random process and are easier to handle than the JDF.

- **Mean**: The mean of the k-th sample of a random process $\mathcal{Y}$ is defined as:

$$\mu_{\mathcal{Y}}(k) = \mathbb{E}[Y_k] = \int_{-\infty}^{\infty} y \, f_{Y_k}(y) \, dy$$

  where $\mathbb{E}[\cdot]$ denotes the expectation operator.

- **Variance**: The variance of the k-th sample of the random process $\mathcal{Y}$ is defined as:

$$\sigma^2_{\mathcal{Y}}(k) = \mathrm{Var}(Y_k) = \mathbb{E}[(Y_k - \mu_{\mathcal{Y}}(k))^2] = \int_{-\infty}^{\infty} (y - \mu_{\mathcal{Y}}(k))^2 \, f_{Y_k}(y) \, dy.$$

  If the variance is constant over time, the process is said to be **homoskedastic**; otherwise, it is called **heteroskedastic**.

- **(Auto)covariance**: The autocovariance between the k-th sample of a random process $\mathcal{Y}$ and the sample lagged by h sampling periods (i.e., a lag of $h \cdot T$ time units) is defined as:

$$C_{\mathcal{Y}}(k, h) = \mathrm{Cov}(Y_k, Y_{k-h}) = \mathbb{E}[(Y_k - \mu_{\mathcal{Y}}(k))(Y_{k-h} - \mu_{\mathcal{Y}}(k-h))].$$

  This covariance is used to measure the linear dependence (or lack thereof) between two samples in a stochastic process. Note that even when $C_{\mathcal{Y}}(k, h) = 0$, there can still be nonlinear dependence between Y_k and Y_{k-h}. Furthermore, for $h = 0$, the autocovariance reduces to the variance, $C_{\mathcal{Y}}(k, 0) = \sigma^2_{\mathcal{Y}}(k)$.

- **(Auto)correlation**: The autocorrelation function (ACF) is a normalized version of the autocovariance, defined as:

$$R_{\mathcal{Y}}(k, h) = \frac{\mathrm{Cov}(Y_k, Y_{k-h})}{\sigma_{\mathcal{Y}}(k)\sigma_{\mathcal{Y}}(k-h)} \in [-1, 1].$$

The autocorrelation assesses how the current value Y_k of a random process is correlated with its past values.

A random process $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_L\}$ is said to be **Gaussian** if the vector of random variables $[Y_1, Y_2, \ldots, Y_L]^\intercal$ follows a multivariate jointly Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu}_\mathcal{Y} = [\mathbb{E}[Y_1], \mathbb{E}[Y_2], \ldots, \mathbb{E}[Y_L]]^\intercal \in \mathbb{R}^L$ and autocovariance matrix $\Sigma_\mathcal{Y} \in \mathbb{R}^{L \times L}$. The entries of the covariance matrix $\Sigma_\mathcal{Y}$ are given by the autocovariance function, such that the (i, j)-th entry satisfies: $[\Sigma_\mathcal{Y}]_{i,j} = \text{Cov}(Y_i, Y_j)$. One key implication of a random process being Gaussian is that the entire process is fully characterized by its mean vector $\boldsymbol{\mu}_\mathcal{Y}(k)$ and the autocovariance matrix $\Sigma_\mathcal{Y}$. This implies that knowing these two parameters is sufficient to describe the statistical properties of the process. Moreover, any linear combination of the components of a Gaussian process is also Gaussian. This property simplifies the analysis and modeling of such processes, particularly in time series and signal processing, where many complex behaviors can be reduced to operations involving the mean and covariance. Furthermore, for a Gaussian process, if the covariance between two random variables is zero, the variables are not only uncorrelated but also statistically independent, a property that does not necessarily hold for non-Gaussian processes.

Example 1: A Simple Random Process

Consider a random process $\mathcal{Y} = \{Y_k : k \in \mathbb{N}\}$ defined by the following recursion:

$$Y_k = \phi Y_{k-1} + \epsilon_k \text{ for all } k \in \mathbb{N}, \text{ with deterministic initial condition } Y_0 = 0,$$

where $|\phi| < 1$ is a constant parameter, and ϵ_k is a sequence of independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance σ_ϵ^2. This process is an example of an AR(1) process (to be covered in coming chapters) and its statistical properties are as follows:

- **Mean**: Since ϵ_k has zero mean, the mean of the process $\mathcal{Y}$ is given by:

  $$\mu_\mathcal{Y}(k) = \mathbb{E}[Y_k] = \phi \mathbb{E}[Y_{k-1}] + \mathbb{E}[\epsilon_k].$$

  Since $Y_0 = 0 = \mathbb{E}[Y_0]$ and $\mathbb{E}[\epsilon_k] = 0$ for all k, it follows that $\mu_\mathcal{Y}(k) = 0$ for all k.

- **Variance**: The variance of the k-th sample is:

  $$\sigma_\mathcal{Y}^2(k) = \text{Var}(Y_k) = \text{Var}(\phi Y_{k-1} + \epsilon_k).$$

  Given that Y_{k-1} and ϵ_k are independent, we have:

  $$\sigma_\mathcal{Y}^2(k) = \phi^2 \text{Var}(Y_{k-1}) + \sigma_\epsilon^2 = \phi^2 \sigma_\mathcal{Y}^2(k-1) + \sigma_\epsilon^2.$$

  Since $Y_0 = 0 = \sigma_\mathcal{Y}^2(0)$, the solution to this recursion is (exercise):

  $$\sigma_\mathcal{Y}^2(k) = \sigma_\epsilon^2 \sum_{i=0}^{k-1} \phi^{2i} = \sigma_\epsilon^2 \frac{1 - \phi^{2k}}{1 - \phi^2}.$$

  Therefore, in the limit $k \to \infty$, we have that $\sigma_\mathcal{Y}^2(k) \to \frac{\sigma_\epsilon^2}{1 - \phi^2}$.

- **95% Confidence Interval**: Since the noise terms ϵ_k are Gaussian and independent, the random variable Y_k is a linear combination of independent Gaussian random variables. Thus, by the properties of linear combinations of Gaussian variables, Y_k is also Gaussian for all k. Therefore, a 95% confidence interval for Y_k can be constructed as $\mu_{\mathcal{Y}}(k) \pm 1.96 \cdot \sigma_{\mathcal{Y}}(k)$. Since $\mu_{\mathcal{Y}}(k) = 0$, the 95% confidence interval simplifies to:

$$Y_k \in \left[-1.96 \cdot \sigma_\epsilon^2 \frac{1-\phi^{2k}}{1-\phi^2}, 1.96 \cdot \sigma_\epsilon^2 \frac{1-\phi^{2k}}{1-\phi^2} \right].$$

  This interval represents the range in which the true value of Y_k will lie with 95% probability, given the known distribution of Y_k.

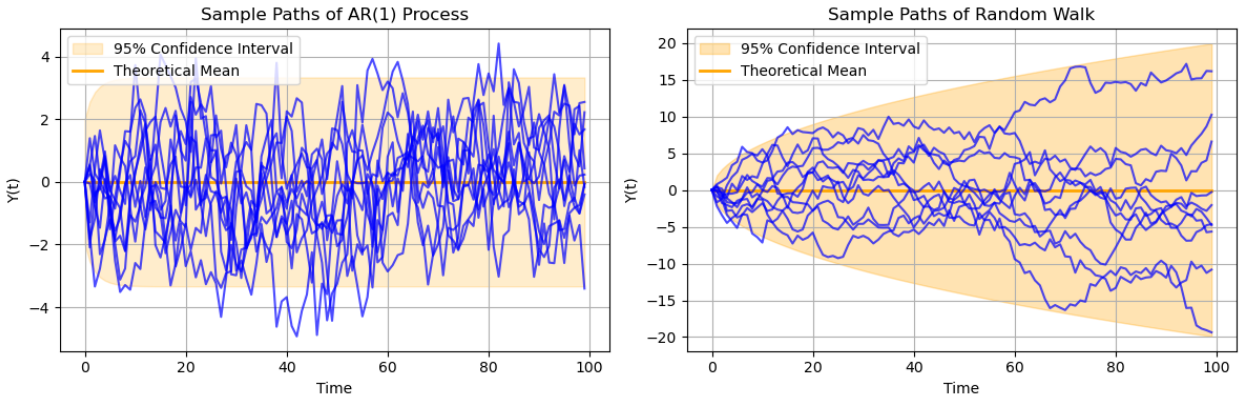

Figure 3: Comparison of the sample paths of an AR(1) process (left) and a random walk with zero drift (right), both starting from $Y_0 = 0$. The shaded regions represent the 95% confidence intervals based on the theoretical variance, i.e., $[\mu_{\mathcal{Y}}(k) \pm 2 \cdot \sigma_{\mathcal{Y}}(k)]$, for each process. The theoretical mean is plotted in orange.

Example 2: Random Walk with Drift

Consider a random process $\mathcal{Y} = \{Y_k : k \in \mathbb{N}\}$ defined by the following recursion:

$$Y_k = \delta + Y_{k-1} + \epsilon_k \text{ for all } k \in \mathbb{N}, \text{ with initial condition } Y_0 = 0,$$

where δ is a constant called the *drift*, and ϵ_k is a sequence of i.i.d. Gaussian random variables with zero mean and variance σ_ϵ^2. This process is an example of a **random walk with drift** and its statistical properties are as follows:

- **Mean**: The mean of the process $\mathcal{Y}$ is given by:

$$\mu_{\mathcal{Y}}(k) = \mathbb{E}[Y_k] = \mathbb{E}[\delta + Y_{k-1} + \epsilon_k] = \delta + \mathbb{E}[Y_{k-1}] + \mathbb{E}[\epsilon_k].$$

  Since $\mathbb{E}[\epsilon_k] = 0$, we have $\mu_{\mathcal{Y}}(k) = \delta + \mu_{\mathcal{Y}}(k-1)$. Since $Y_0 = 0$, this recurrence relation gives:

$$\mu_{\mathcal{Y}}(k) = k\delta.$$

  This indicates that the mean of the process increases linearly over time, reflecting the effect of the drift δ.

- **Variance**: The variance of the k-th sample is:

$$\sigma^2_{\mathcal{Y}}(k) = \text{Var}(Y_k) = \text{Var}(\delta + Y_{k-1} + \epsilon_k).$$

Given that Y_{k-1} and ϵ_k are independent (and therefore uncorrelated), we have:

$$\sigma^2_{\mathcal{Y}}(k) = \text{Var}(Y_{k-1}) + \text{Var}(\epsilon_k) = \sigma^2_{\mathcal{Y}}(k-1) + \sigma^2_\epsilon.$$

Since $Y_0 = 0$, the variance evolves as:

$$\sigma^2_{\mathcal{Y}}(k) = k\sigma^2_\epsilon.$$

Thus, the variance of the process increases linearly with time (and the standard deviation as the square root of time), indicating that the process becomes more variable as time progresses.

- **95% Confidence Interval**: Because ϵ_k are Gaussian random variables, Y_k, as a sum of Gaussian random variables, is also Gaussian at each time k. Using this distribution, we can compute a 95% confidence interval for Y_k as:

$$Y_k \in [\mu_{\mathcal{Y}}(k) - 1.96 \cdot \sigma_{\mathcal{Y}}(k), \mu_{\mathcal{Y}}(k) + 1.96 \cdot \sigma_{\mathcal{Y}}(k)].$$

Since $\mu_{\mathcal{Y}}(k) = k\delta$ and $\sigma_{\mathcal{Y}}(k) = \sqrt{k}\sigma_\epsilon$, the confidence interval becomes:

$$Y_k \in [k\delta - 1.96 \cdot \sqrt{k}\sigma_\epsilon, k\delta + 1.96 \cdot \sqrt{k}\sigma_\epsilon].$$

This interval gives a range in which we expect Y_k to fall with 95% probability at each time step k, accounting for both the drift δ and the increasing variance over time.

2.2 Stationarity in Time Series

Stationarity is a fundamental concept in time series analysis that refers to the idea that the statistical properties of a time series do not change over time. When a process is stationary, it is easier to model and make predictions because its behavior is consistent and predictable over time.

2.2.1 Strong-Sense Stationarity (SSS)

One of the strongest forms of stationarity is known as **strong-sense stationarity (SSS)**. A random process $\mathcal{Y} = \{Y_k : k \in \mathbb{N}\}$ is SSS if the joint distribution of any finite collection of random variables from the process is invariant under shifts in time. Specifically, for any collection of discrete time indices $k_1, k_2, \ldots, k_L$, the joint distribution of the corresponding random variables $(Y_{k_1}, Y_{k_2}, \ldots, Y_{k_L})$ remains unchanged if we shift all time indices by a constant h. Formally, the process is SSS if:

$$\Pr(Y_{k_1} \leq y_1, Y_{k_2} \leq y_2, \ldots, Y_{k_L} \leq y_L) = \Pr(Y_{k_1+h} \leq y_1, Y_{k_2+h} \leq y_2, \ldots, Y_{k_L+h} \leq y_L)$$

for all $L \in \mathbb{N}$, $h \in \mathbb{N}$, and any values $y_1, y_2, \ldots, y_L \in \mathbb{R}$. In other words, the entire joint probability structure of the process does not change over time.

The condition of strong-sense stationarity is very stringent and has several important implications for the random process $\mathcal{Y}$. In particular:

- **Invariance of Means and Variances:** Since the entire joint distribution is invariant under time shifts, this implies that the mean and variance of the process must be constant over time. Then:
$$\mu_{\mathcal{Y}}(k) = \mu_{\mathcal{Y}}(k+h) = \mu_{\mathcal{Y}} \text{ and } \sigma_{\mathcal{Y}}(k) = \sigma_{\mathcal{Y}}(k+h) = \sigma_{\mathcal{Y}},$$
for all $k, h \in \mathbb{N}$. This indicates that the expected value of the process does not change as time progresses.
- **Time-Invariance of the Covariances and Autocorrelation:** Strong-sense stationarity implies that the covariance and the correlation between two random variables Y_k and Y_{k+h}, depends only on the time difference h, not on the specific times k or $k+h$. This means that the pair-wise statistical dependencies within the process are stationary and do not change over time.

Example 3: White Noise Process

A common example of a strongly stationary process is the **white noise process**. In this case, the random variables Y_k are independent and identically distributed (i.i.d.) with zero mean and constant variance σ^2. Since each Y_k is independent of all others and has the same joint distribution regardless of time, the white noise process trivially satisfies the condition for strong-sense stationarity.

- **Mean:** The mean of the white noise process is constant and given by:
$$\mu_{\mathcal{Y}}(k) = \mathbb{E}[Y_k] = 0 \text{ for all } k.$$
This reflects the fact that, on average, the random variables in a white noise process have a value of zero.
- **Covariance:** Due to the independence of the random variables in the white noise process, the covariance is zero for any non-zero lag h:
$$C_{\mathcal{Y}}(k,h) = \begin{cases} \sigma^2 & \text{if } h = 0, \\ 0 & \text{if } h \neq 0, \end{cases}$$
where σ^2 is the constant variance of the process. This means that the random variables are uncorrelated unless they coincide in time.
- **Autocorrelation:** Given the form of the covariance function, the autocorrelation function is:
$$R_{\mathcal{Y}}(h) = \begin{cases} 1 & \text{if } h = 0, \\ 0 & \text{if } h \neq 0. \end{cases}$$
This indicates that each random variable in the white noise process is only correlated with itself and uncorrelated with all other variables, regardless of the time lag.

In summary, the white noise process is a simple yet fundamental example of a strongly stationary process, characterized by zero mean, constant variance, and no autocorrelation between different time points.

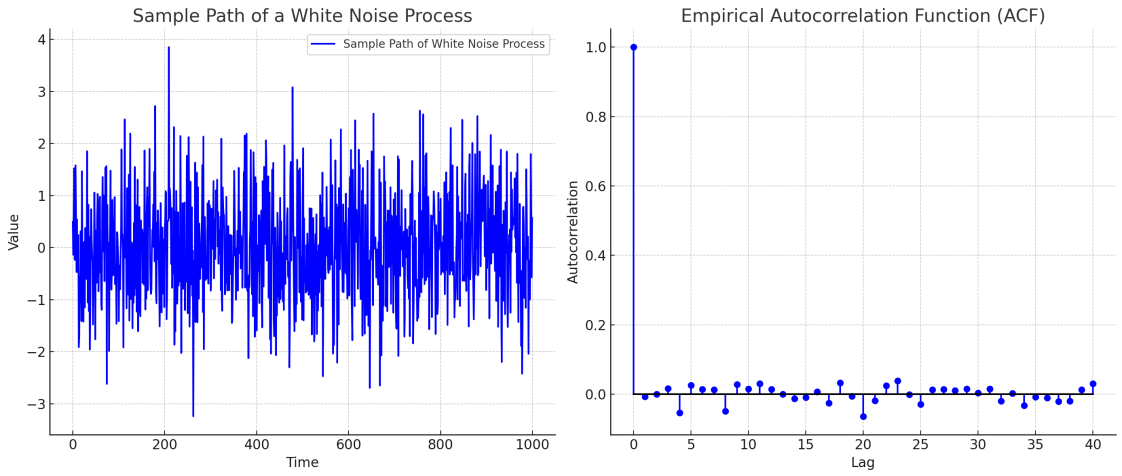

Figure 4: (Left) A typical sample path of a white noise and (right) its empirical ACF.

Strong-sense stationarity plays a critical role in many areas of time series analysis and stochastic processes. Stationarity simplifies modeling because it ensures that the statistical properties of the process remain constant over time. However, due to its stringent requirements, SSS is often difficult to achieve in practice. Many real-world processes, such as stock prices or environmental data, exhibit time-varying trends or volatility, making them non-stationary.

2.3 Weak-Sense Stationarity (WSS)

While strong-sense stationarity (**SSS**) requires that the entire joint distribution of the process be invariant under time shifts, weak-sense stationarity (**WSS**) is a less restrictive condition. **WSS** requires only that the first and second moments (mean, variance, and covariance) be constant over time, which is often sufficient for practical applications. Since **WSS** does not impose conditions on the higher-order moments or the full distribution, it is easier to achieve and more commonly encountered in real-world processes.

A random process $\mathcal{Y} = \{Y_k : k \in \mathbb{N}\}$ is said to be **Weak-Sense Stationary** (**WSS**) if it satisfies the following three conditions:

1. **Constant Mean**: The mean of the process, $\mu_{\mathcal{Y}}(k) = \mathbb{E}[Y_k]$, must be constant for all k, i.e., $\mu_{\mathcal{Y}}(k) = \mu$ for all k. This means the expected value of the process does not change over time.

2. **Constant Variance**: The variance of each random variable in the process, $\sigma^2_{\mathcal{Y}}(k) = \mathrm{Var}(Y_k)$, must also be constant for all k, i.e., $\sigma^2_{\mathcal{Y}}(k) = \sigma^2$ for all k. This indicates that the variability of the process remains constant over time.

3. **Time-Invariant Autocovariance**: The autocovariance between two random variables Y_k and Y_{k+h}, denoted as $C_{\mathcal{Y}}(k, h) = \mathrm{Cov}(Y_k, Y_{k+h})$, depends only on the time difference (or lag) h, and not on the specific times k or $k + h$. Thus, we have: $C_{\mathcal{Y}}(k, h) = C_{\mathcal{Y}}(h)$ for all k, h. This implies that the covariance between two points in the process is determined only by the lag h, not by their absolute positions in the time series.

Weak-sense stationarity is fundamental in time series analysis because it simplifies the modeling and analysis of stochastic processes. Many time series models, such as autoregressive and moving average models, rely on the assumption of stationarity. Stationarity ensures that the statistical properties of the process—specifically, the mean, variance, and autocovariance structure—remain constant over time. This stability allows for the derivation of reliable forecasts and valid inferential procedures.

Before applying models that assume stationarity, it is important to determine whether the observed sample path is likely to be a realization of a stationary stochastic process. Empirical verification of stationarity can be carried out using several tools, detailed below.

1. **Basic Statistical Tests**: Stationarity implies that the mean, variance, and autocovariance of the time series do not vary with time. A preliminary assessment of stationarity can be performed using the following techniques:

   - **Plot the Time Series**: Visually inspect the time series to check whether it fluctuates around a constant mean and exhibits consistent variability. If clear trends or changes in variance are observed, the process is likely non-stationary.
   - **Rolling Statistics**: Compute and plot the rolling mean and rolling variance using a fixed window size. If the rolling statistics remain approximately constant, this is indicative of stationarity. Significant shifts in these statistics over time suggest non-stationarity.
   - **Autocorrelation Function (ACF)**: Plot the autocorrelation function to assess how the autocorrelations evolve with increasing lags. For a stationary process, the autocorrelation function typically decays towards zero relatively quickly. A persistent autocorrelation at higher lags may signal non-stationarity.

   While these methods provide initial insights, formal statistical tests are required to confirm stationarity.

2. **Augmented Dickey-Fuller (ADF) Test**: The ADF test is a commonly employed hypothesis test to formally assess the stationarity of a time series. The null and alternative hypotheses for the ADF test are:

   $$H_0 : \gamma = 0 \quad \text{(the series is non-stationary) vs. } H_A : \gamma < 0 \quad \text{(the series is stationary).}$$

   If the test statistic is sufficiently negative, we reject H_0, suggesting that the time series is stationary. The ADF test provides a p-value, and if the p-value is below a chosen significance level (typically 0.05), we reject the null hypothesis of non-stationarity. This test is widely used in practice and is implemented in the method `adfuller` in the Python package `statsmodels`. The function returns both the test statistic and the p-value, along with additional information, enabling users to make a formal decision on stationarity based on the test results.

Example 4: Autocovariance of a Random Walk

Consider a random walk process defined by:

$$Y_k = Y_{k-1} + \epsilon_k, \quad \epsilon_k \sim \mathcal{N}(0, \sigma_\epsilon^2) \text{ with initial condition } Y_0 = 0,$$

where ϵ_k are independent and identically distributed (i.i.d.) Gaussian random variables with

zero mean and constant variance σ_ϵ^2. For this random walk, the mean is zero for all k because the increments ϵ_k have zero mean. Therefore, the covariance becomes $C_{\mathcal{Y}}(h) = \mathbb{E}[Y_k Y_{k-h}]$. We can express Y_k and Y_{k-h} as a sum of the increments ϵ_k:

$$Y_k = Y_0 + \sum_{i=1}^{k} \epsilon_i \text{ and } Y_{k-h} = Y_0 + \sum_{i=1}^{k-h} \epsilon_i.$$

Now, we compute the covariance:

$$C_{\mathcal{Y}}(h) = \mathbb{E}\left[\left(\sum_{i=1}^{k} \epsilon_i\right)\left(\sum_{j=1}^{k-h} \epsilon_j\right)\right].$$

Expanding this expression, we get:

$$C_{\mathcal{Y}}(h) = \sum_{i=1}^{k}\sum_{j=1}^{k-h} \mathbb{E}[\epsilon_i \epsilon_j].$$

Since ϵ_i are independent, $\mathbb{E}[\epsilon_i \epsilon_j] = 0$ for $i \neq j$, and for $i = j$, we have $\mathbb{E}[\epsilon_i^2] = \sigma_\epsilon^2$. Therefore, the only terms that contribute to the sum are those with $i = j$, yielding:

$$C_{\mathcal{Y}}(h) = \sum_{i=1}^{k-h} \sigma_\epsilon^2 = (k-h)\sigma_\epsilon^2.$$

Thus, the covariance of a random walk is:

$$C_{\mathcal{Y}}(h) = (k-h)\sigma_\epsilon^2 \quad \text{for } h \leq k.$$

This shows that the covariance depends linearly on k, the current time, and decreases with increasing lag h; hence, a random walk is *not* stationary.

Example 5: Autocovariance of an Autoregressive Process AR(1)

Consider the AR(1) stochastic process, defined by the recursion:

$$Y_k = \phi Y_{k-1} + \epsilon_k \quad \text{with} \quad |\phi| < 1,$$

where ϵ_k is a white noise process with zero mean and variance σ_ϵ^2. As we cab observe in Fig. 5-Left, the process is not be stationary for small k, since the deterministic initial condition induces at initial growth of the variance. However, for large k, the variance and autocorrelation of the AR(1) process becomes weak-sense stationary exponentially fast as the influence of the initial condition decays over time. The analysis of the mean and the variance in Example 1 already established that the mean is zero and the variance converges to the constant $\frac{\sigma_\epsilon^2}{1-\phi^2}$ as $k \to \infty$.
Let us now make an analysis of the autocovariance and the autocorrelation for k large. To derive the autocovariance, we start by writing Y_{k+h} in terms of Y_k using the recursion:

$$Y_{k+h} = \phi^h Y_k + \phi^{h-1}\epsilon_{k+1} + \phi^{h-2}\epsilon_{k+2} + \cdots + \epsilon_{k+h}.$$

Multiplying this expression by Y_k and taking expectations, we get:

$$\mathbb{E}[Y_k Y_{k+h}] = \mathbb{E}\left[Y_k\left(\phi^h Y_k + \sum_{i=1}^{h} \phi^{h-i}\epsilon_{k+i}\right)\right].$$

Using the fact that Y_k and ϵ_{k+i} are uncorrelated for all i, the cross terms vanish, and we are left with:

$$\mathbb{E}[Y_k Y_{k+h}] = \phi^h \mathbb{E}[Y_k^2].$$

Since the variance of Y_k is constant and given by $\sigma_{\mathcal{Y}}^2 = \mathbb{E}[Y_k^2] = \frac{\sigma_\epsilon^2}{1-\phi^2}$, we substitute this into the equation:

$$C_{\mathcal{Y}}(h = \mathbb{E}[Y_k Y_{k+h}] = \phi^h \frac{\sigma_\epsilon^2}{1-\phi^2}.$$

Autocorrelation: The autocorrelation function is the normalized autocovariance. Since $C_{\mathcal{Y}}(0) = \frac{\sigma_\epsilon^2}{1-\phi^2}$, the autocorrelation function is given by:

$$R_{\mathcal{Y}}(h) = \begin{cases} 1 & \text{if } h = 0, \\ \phi^{|h|} & \text{if } h \neq 0. \end{cases}$$

For large lags h, if $|\phi| < 1$, the autocorrelation *decays exponentially*, and $R_{\mathcal{Y}}(h) \to 0$ as $h \to \infty$. This reflects that the AR(1) process becomes progressively uncorrelated for distant observations. In Fig. 5 we include a realization of a sample path of length $L = 1,000$, including the evolution of the empirical mean and variance using a rolling window. In the right subplot, we also include the empirical and theoretical autocorrelation function.

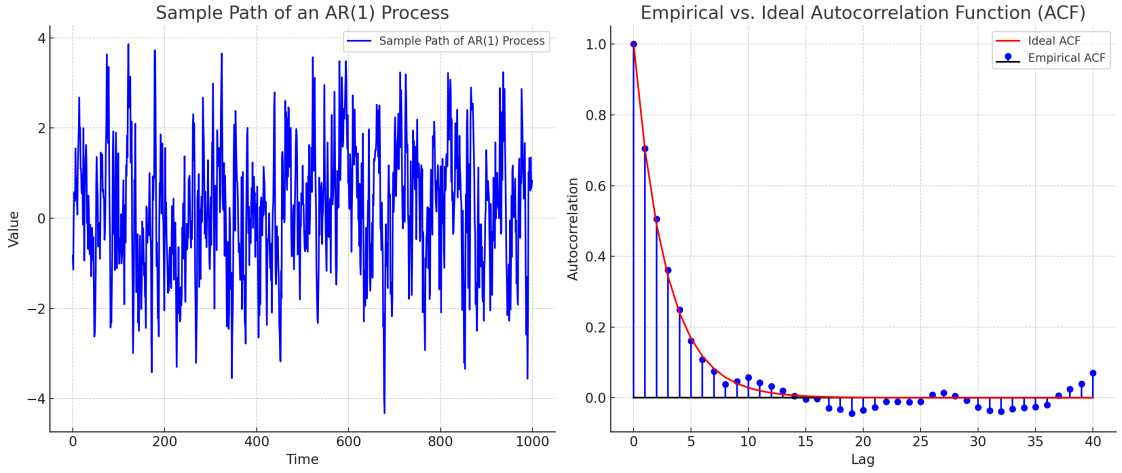

Figure 5: (Left) A typical sample path of a white noise and (right) its empirical ACF.

3 Multivariate Time Series

A **multivariate time series** involves observing and analyzing multiple interrelated time-dependent variables simultaneously. Unlike a univariate time series, which focuses on a single variable, multi-

variate time series account for several variables together, enabling the study of potential interactions and dependencies among them. Let us denote a discrete-time multivariate time series as $\boldsymbol{\mathcal{Y}}$, where $\mathbf{Y}(k) = [Y_1(k), Y_2(k), \ldots, Y_k(k)]^\intercal \in \mathbb{R}^k$ represent a k-dimensional vector with $Y_i(k)$ denoting the value of the i-th variable at time slot k. The strength of multivariate analysis lies in its ability to capture dynamic interdependencies between variables over time. This is particularly valuable in fields such as economics, finance, and environmental science, where the behavior of one variable may be influenced by or dependent on others.

3.1 Statistical Properties of Multivariate Time Series

To analyze multivariate time series effectively, a variety of statistical tools are employed, each designed to capture the relationships between the individual series and their temporal dynamics. Below, we provide a detailed overview of these key statistical measures.

- **Mean Vector:** The **mean vector** $\boldsymbol{\mu}(k)$ provides a summary of the expected values for each component of the multivariate time series at a given time t. Mathematically, it is expressed as:
$$\boldsymbol{\mu}(k) = \mathbb{E}[\mathbf{Y}(k)] = \begin{bmatrix}\mathbb{E}[Y_1(k)] & \mathbb{E}[Y_2(k)] & \ldots & \mathbb{E}[Y_k(k)]\end{bmatrix}^\intercal \in \mathbb{R}^k.$$
This vector encapsulates the average behavior of each component, offering a foundational perspective on the trends present in the data.

- **Covariance Matrix:** The **covariance matrix** $\Sigma(t_1, t_2)$ quantifies the degree to which the components of the time series co-vary across two time points t_1 and t_2. It is defined as:
$$\Sigma(t_1, t_2) = \mathbb{E}[(\mathbf{Y}(t_1) - \boldsymbol{\mu}(t_1))(\mathbf{Y}(t_2) - \boldsymbol{\mu}(t_2))^\intercal] \in \mathbb{R}^{k\times k}.$$
The covariance matrix provides insights into both the variability of each individual series (represented by the diagonal elements) and the co-movement between different series (represented by the off-diagonal elements). For WSS multivariate time series, where statistical properties remain constant over time, the covariance matrix depends only on the lag $h = t_2 - t_1$, simplifying to:
$$\Sigma(h) = \mathbb{E}[(\mathbf{Y}(k) - \boldsymbol{\mu})(\mathbf{Y}(k-h) - \boldsymbol{\mu})^\intercal] \in \mathbb{R}^{k\times k}.$$
In this case, $\Sigma(h)$ captures how the components are related across different time lags, with variances along the diagonal and covariances in the off-diagonal.

- **Correlation Coefficient:** While covariance indicates the direction and magnitude of how two variables vary together, the **correlation coefficient** standardizes this relationship by accounting for the variance of each variable. The Pearson correlation coefficient between two signals Y_i and Y_j at time t is defined as:
$$\rho(Y_i(k), Y_j(k)) = \frac{\mathrm{Cov}(Y_i(k), Y_j(k))}{\sqrt{\mathrm{Var}(Y_i(k))\mathrm{Var}(Y_j(k))}} \in [-1, 1].$$
Correlation coefficients provide a scale-independent measure of association, making it easier to interpret the strength and direction of the relationships between variables in a multivariate time series.

- **Cross-Covariance Function:** The **cross-covariance function** $\text{Cov}(Y_i(k), Y_j(k-h))$ measures the covariance between two signals $Y_i(k)$ and $Y_j(k-h)$ at different lags h. It is defined as:

$$\text{Cov}(Y_i(k), Y_j(k-h)) = \mathbb{E}[(Y_i(k) - \mu_{Y_i}(k))(Y_j(k-h) - \mu_{Y_j}(k-h))].$$

  This function provides insights into the linear dependence between two variables across different time lags.

- **Cross-Correlation Function:** The **cross-correlation function** (CCF) $\text{CCF}(Y_i, Y_j, h)$ measures the temporal correlation between two signals $Y_i(k)$ and $Y_j(k)$ at different lags h. It is defined as:

$$\text{CCF}(Y_i, Y_j, h) = \frac{\text{Cov}(Y_i(k), Y_j(k-h))}{\sqrt{\text{Var}(Y_i(k))\text{Var}(Y_j(k-h))}}.$$

  The cross-correlation function normalizes the covariance by the standard deviations of the variables, providing a scale-independent measure of association. It helps determine whether changes in one variable lead or lag changes in another, revealing the direction and strength of temporal dependencies. Positive cross-correlation values indicate that increases in Y_i are associated with increases in Y_j at a given lag, while negative values suggest an inverse relationship. The CCF is particularly useful in identifying leading and lagging relationships between variables, where one series may consistently lead or follow another over a specific time interval. This information is critical in fields like finance, where the behavior of one asset might predict the movement of another.

3.2 Python Example: Multivariate Statistical Measures

In this example, we compute key statistical measures for a multivariate time series using Python. We will simulate a multivariate time series with three interrelated variables, compute the mean vector, covariance matrix, and visualize the relationships using scatter plots and cross-correlation plots.

1. **Simulating Time Series Data:** In this first code cell, we generate a time series using the multivariate normal distribution. This step simulates three variables with predefined covariance.

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from numpy.random import multivariate_normal
from scipy.stats import zscore

# Simulate multivariate time series
np.random.seed(42)
n = 200  # number of time points
mean = [0, 0, 0]  # mean vector for the series

# Covariance matrix for 3 interrelated time series
cov = [[1, 0.8, 0.5],
       [0.8, 1, 0.4],
       [0.5, 0.4, 1]]

# Generating multivariate time series data
data = multivariate_normal(mean, cov, n)
```

```
time = np.arange(n)

# Creating a DataFrame
df = pd.DataFrame(data, columns=['Y1', 'Y2', 'Y3'])
df['time'] = time

# Display the first few rows
df.head()
```

In this cell, we simulate 200 time points for three variables, stored in the DataFrame `df`. The covariance matrix is predefined to introduce relationships between the variables.

2. **Mean Vector:** We compute the mean vector, which summarizes the average behavior of each time series component.

```
# Mean vector
mean_vector = df[['Y1', 'Y2', 'Y3']].mean()
print("Mean Vector:")
print(mean_vector)
```

```
Mean Vector:  Y1:  -0.073002, Y2:  -0.120167, Y3:  0.056289, dtype:  float64
```

This step outputs the mean values for each of the three time series components, providing insights into their central tendency.

3. **Covariance Matrix:** The (empirical) covariance matrix is computed to quantify how much each pair of variables co-varies.

```
# Covariance matrix
cov_matrix = df[['Y1', 'Y2', 'Y3']].cov()
print("Covariance Matrix:")
print(cov_matrix)
```

	Y1	Y2	Y3
Y1	0.838843	0.614269	0.407596
Y2	0.614269	0.809994	0.256549
Y3	0.407596	0.256549	0.923854

The empirical covariance matrix helps to understand the relationships between the time series components, with variances on the diagonal and covariances in the off-diagonal. The difference between the theoretical covariance matrix and the empirical estimation above is due to the fact that we are working with a finite sample of data and random variation affects the estimates of covariance when the sample size is finite.

4. **Scatter Plots and Correlation Coefficients:** Scatter plots are used to visualize the pairwise relationships between the variables. The correlation coefficients are computed and displayed in the titles of the plots.

```
import matplotlib.pyplot as plt

# Scatter plots between Y1, Y2, and Y3
plt.figure(figsize=(10, 4))

# Compute the Pearson correlation coefficients
corr_Y1_Y2 = df['Y1'].corr(df['Y2'])
corr_Y1_Y3 = df['Y1'].corr(df['Y3'])
```

```
corr_Y2_Y3 = df['Y2'].corr(df['Y3'])

# Y1 vs Y2
plt.subplot(1, 3, 1)
plt.scatter(df['Y1'], df['Y2'], alpha=0.5)
plt.title(f'Y1 vs Y2\nCorr: {corr_Y1_Y2:.2f}')
plt.xlabel('Y1')
plt.ylabel('Y2')

# Y1 vs Y3
plt.subplot(1, 3, 2)
plt.scatter(df['Y1'], df['Y3'], alpha=0.5)
plt.title(f'Y1 vs Y3\nCorr: {corr_Y1_Y3:.2f}')
plt.xlabel('Y1')
plt.ylabel('Y3')

# Y2 vs Y3
plt.subplot(1, 3, 3)
plt.scatter(df['Y2'], df['Y3'], alpha=0.5)
plt.title(f'Y2 vs Y3\nCorr: {corr_Y2_Y3:.2f}')
plt.xlabel('Y2')
plt.ylabel('Y3')

plt.tight_layout()
plt.show()
```

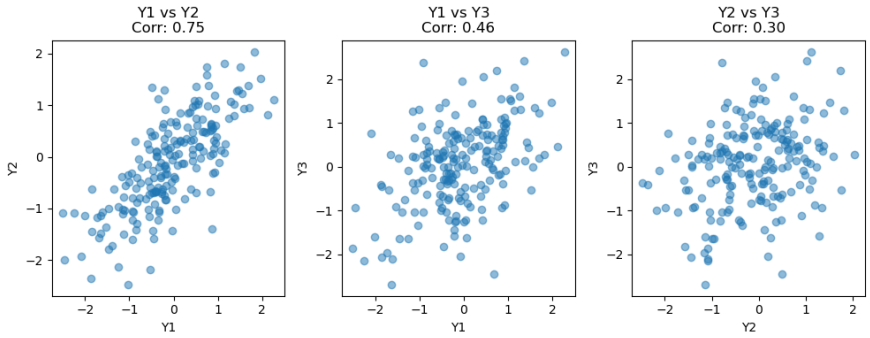

Figure 6: Scatter plots showing relationships between Y_1, Y_2, and Y_3.

5. **Cross-Correlation Function (CCF):** Finally, we compute and plot the cross-correlation functions for each pair of variables across different time lags using a matrix plot.

```
# Standardizing the series (z-scores)
df['Y1_z'] = zscore(df['Y1'])
df['Y2_z'] = zscore(df['Y2'])
df['Y3_z'] = zscore(df['Y3'])

# List of standardized variables
variables = ['Y1_z', 'Y2_z', 'Y3_z']

# Create subplots for cross-correlation matrix
fig, axes = plt.subplots(nrows=3, ncols=3, figsize=(12, 12))
max_lags = 20  # Maximum number of lags

# Plot cross-correlations for each pair of variables
```

```
for i, var1 in enumerate(variables):
        for j, var2 in enumerate(variables):
                if i == j:
                        # Autocorrelation on the diagonal
                        axes[i, j].xcorr(df[var1], df[var2], maxlags=
                            max_lags)
                        axes[i, j].set_title(f'Autocorrelation: {var1}')
                else:
                        # Cross-correlation for off-diagonal elements
                        axes[i, j].xcorr(df[var1], df[var2], maxlags=
                            max_lags)
                        axes[i, j].set_title(f'Cross-correlation: {var1} &
                            {var2}')

                axes[i, j].set_xlabel('Lag')
                axes[i, j].set_ylabel('Cross-correlation')

plt.tight_layout()
plt.show()
```

This cell produces a matrix of plots that show both the autocorrelations (diagonal) and cross-correlations (off-diagonal) between the variables for different time lags.

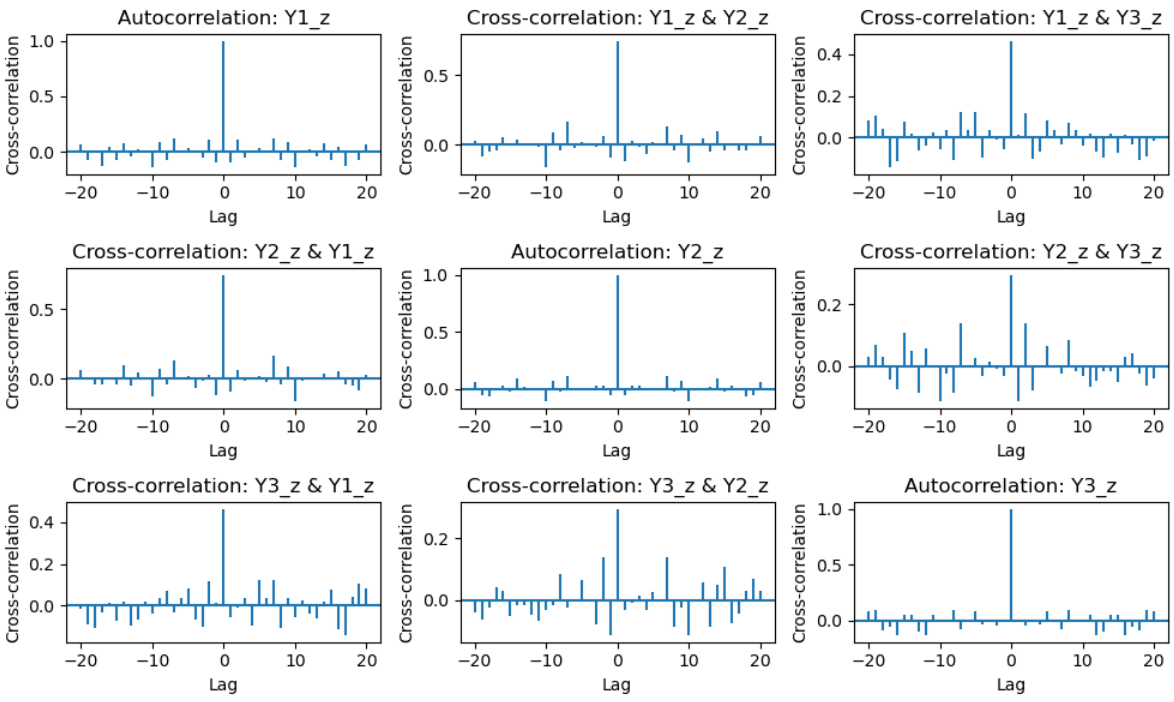

Figure 7: Cross-correlation matrix for Y_1, Y_2, and Y_3.

The cross-correlation function helps identify the lead-lag relationships between variables across different time lags, revealing important temporal dependencies.