Stationarity and Ergodicity in Time Series

From Intuition to Theory

Diogo Ribeiro

Data Science and Applied Mathematics

November 11, 2025

Outline

Why Time Series?

- Many real-world phenomena are naturally ordered in time:
  - Financial returns, volatility, prices.
  - Temperature and climate indicators.
  - Sensor data, biomedical signals, internet traffic.
- We often observe **one long history**, not repeated experiments.
- Aim:
  - Understand the probabilistic mechanism generating the series.
  - Forecast future values.
  - Detect changes, regimes, or anomalies.

Stationarity and ergodicity tell us when the past is informative about the future and about the underlying process.

Typical Time Series Questions

- Is the mean level of the series stable over time?
- Does the variability change (volatility clustering, heteroscedasticity)?
- Are there trends or seasonal patterns?
- Can we treat one observed trajectory as *representative* of the process?

Stationarity addresses stability of distributions over time. **Ergodicity** connects time averages and ensemble averages.

Stochastic Process: Interaction with Time

- A time series is a realization of a stochastic process:

$$\{X_t\}_{t\in\mathbb{Z}} \quad \text{or} \quad \{X_t\}_{t\in\mathbb{R}}.$$

- For each time t, X_t is a random variable defined on some probability space.
- A realization (sample path) is one function $t \mapsto x_t$ generated by the process.
- In practice we observe:

$$x_1, x_2, \ldots, x_T.$$

The challenge: infer properties of the *process* from a single time-ordered sample.

Two Perspectives: Time and Ensemble

- **Time perspective** (within one realization):
  - Follow X_t for $t = 1, \ldots, T$.
  - Compute time averages, such as sample mean and sample variance.
- **Ensemble perspective** (across realizations):
  - At fixed time t, consider many copies $X_t^{(1)}, \ldots, X_t^{(N)}$.
  - The distribution of these is the *ensemble distribution* at time t.
- In practice, we almost never have many independent realizations from the same process.

This tension between time and ensemble is where stationarity and ergodicity enter.

Fundamental Problem

- We want ensemble quantities, such as $E[X_t]$, $\mathrm{Var}(X_t)$, or joint distributions.
- We only have a single path:

$$x_1, x_2, \ldots, x_T.$$

- Natural estimators:

$$\bar{X}_T = \frac{1}{T} \sum_{t=1}^{T} X_t, \quad s_T^2 = \frac{1}{T-1} \sum_{t=1}^{T} (X_t - \bar{X}_T)^2.$$

- When is $\bar{X}_T$ a good estimator of $E[X_t]$? When does s_T^2 estimate $\mathrm{Var}(X_t)$?

Answering this rigorously requires both stationarity and ergodicity.

Strict (Strong) Stationarity

Definition. A process $\{X_t\}$ is *strictly stationary* if for any $k \in \mathbb{N}$, any times

$$t_1, \ldots, t_k$$

and any integer shift h, the joint distributions satisfy

$$(X_{t_1}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_k+h}).$$

Intuition:

- The probabilistic structure of the process is invariant under time shifts.
- All finite-dimensional distributions are time-homogeneous.

Weak (Covariance) Stationarity

Definition. A process $\{X_t\}$ is *weakly stationary* (or covariance stationary) if:

1. $E[X_t] = \mu$ is constant for all t.
2. $\mathrm{Var}(X_t) = \sigma^2 < \infty$ is constant for all t.
3. $\mathrm{Cov}(X_t, X_{t+h})$ depends only on h, not on t:

$$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}).$$

This is enough for many linear time series models (ARMA, etc.) and for spectral analysis.

Autocovariance and Autocorrelation

For a weakly stationary process:

- Autocovariance function (ACVF):

$$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}), \quad h \in \mathbb{Z}.$$

- Autocorrelation function (ACF):

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

Properties:

- $\gamma(0) = \sigma^2$.
- $\gamma(-h) = \gamma(h)$.
- $\rho(0) = 1$, and $|\rho(h)| \leq 1$ for all h.

Examples of Stationary Processes

Example 1: White noise

- $X_t \sim \text{i.i.d.}(0, \sigma^2)$.
- $E[X_t] = 0$, $\gamma(0) = \sigma^2$, $\gamma(h) = 0$ for $h \neq 0$.
- Strictly and weakly stationary.

Example 2: AR(1) with $|\phi| < 1$

$$\begin{aligned} &X_t = \phi X_{t-1} + \varepsilon_t, \\ &\varepsilon_t \sim \text{i.i.d.}(0, \sigma_\varepsilon^2), \quad |\phi| < 1. \end{aligned}$$

Then:

$$E[X_t] = 0, \quad \gamma(h) = \frac{\sigma_\varepsilon^2}{1 - \phi^2} \phi^{|h|}.$$

This process is weakly stationary and, under mild conditions, strictly stationary.

Non-stationary Example: Random Walk

Random walk:

$$X_t = X_{t-1} + \varepsilon_t,$$
$$\varepsilon_t \sim \text{i.i.d.}(0, \sigma_\varepsilon^2), \quad X_0 \text{ given.}$$

- $E[X_t] = E[X_0]$ (constant mean).
- $\mathrm{Var}(X_t) = \mathrm{Var}(X_0) + t\sigma_\varepsilon^2$ grows with t.
- The variance is not constant $\rightarrow$ not even weakly stationary.

Differencing $\Delta X_t = X_t - X_{t-1} = \varepsilon_t$ recovers a stationary series (white noise).

Strict vs Weak Stationarity

- Strict stationarity $\Rightarrow$ weak stationarity if second moments exist.
- Weak stationarity does not necessarily imply strict stationarity.
- In practice:
  - We rarely can test strict stationarity.
  - Most modelling frameworks assume weak stationarity.

For many Gaussian processes, weak stationarity actually implies strict stationarity, because the distribution is fully determined by mean and covariance.

Time Averages vs Ensemble Averages

For a weakly stationary process $\{X_t\}$ with finite mean μ:

- Ensemble mean:
$$\mu = E[X_t].$$
- Time average (sample mean) over one trajectory:
$$\bar{X}_T = \frac{1}{T} \sum_{t=1}^{T} X_t.$$

Question: Under what conditions does

$$\bar{X}_T \xrightarrow[T\to\infty]{\text{(some sense)}} \mu?$$

This is the core idea behind *ergodicity*.

Ergodicity in the Mean

Definition (informal). A stationary process $\{X_t\}$ is *ergodic in the mean* if

$$\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow[T\to\infty]{\text{a.s. or in prob.}} E[X_t] = \mu.$$

Interpretation:

- A single long realization is enough to estimate the mean.
- The time average along one trajectory converges to the ensemble expectation.

Ergodicity for Higher Moments

We can extend the notion of ergodicity:

- **Ergodic in variance:**

$$\frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X}_T)^2 \xrightarrow[T\to\infty]{} \mathrm{Var}(X_t).$$

- **Ergodic for autocovariances:**

$$\frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h} \xrightarrow[T\to\infty]{} \gamma(h).$$

In practice, we estimate $\gamma(h)$ from one series and implicitly assume ergodicity.

Ergodic Theorem (Very Informal)

Birkhoff's Ergodic Theorem (informal statement):

- Consider a measure-preserving transformation T on a probability space and an integrable function f.
- Under ergodicity of T, the time average along orbits converges (almost surely) to the space average:

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega) \xrightarrow[n\to\infty]{\text{a.s.}} \int f \, d\mathbb{P}.$$

Time averages are legitimate estimators of expectations if the underlying dynamical system is ergodic.

Stationarity vs Ergodicity

- Stationarity:
  - Distributions do not change with time shifts.
  - Structural property of the process.
- Ergodicity:
  - Time averages equal ensemble averages.
  - Relates individual realizations to the overall distribution.

Relationship:

- Ergodicity *implies* stationarity (in appropriate formulations).
- Stationarity alone does not guarantee ergodicity.

Examples: When Is a Process Ergodic?

- Many ARMA processes with $|\phi_i| < 1$ and i.i.d. innovations are stationary and ergodic.
- Gaussian stationary processes often satisfy ergodic properties under mild conditions.
- A process that is a mixture of two stationary regimes but never switches between them can be stationary and non-ergodic.

Informally, ergodicity fails when different realizations live in different "parts" of the state space and never explore the whole distribution.

Checking Stationarity in Practice

- **Visual inspection:**
  - Plot the series, rolling mean, and rolling variance.
  - Look for structural breaks, changing variance, trends, seasonality.
- **Unit root and stationarity tests:**
  - Augmented Dickey-Fuller (ADF): null = unit root (non-stationary).
  - KPSS: null = stationary.
  - Phillips-Perron, DF-GLS, others.
- **ACF and PACF:**
  - Slowly decaying ACF suggests non-stationarity.
  - Sudden drops suggest stationarity (for ARMA-type processes).

Transformations to Achieve Stationarity

- **Detrending:**
  - Remove deterministic trends (linear or nonlinear).
  - Work with residuals rather than raw series.
- **Differencing:**

$$\nabla X_t = X_t - X_{t-1}, \quad \nabla^d X_t = (1 - B)^d X_t.$$

- **Variance-stabilizing transforms:**
  - Log-transform for strictly positive series.
  - Box-Cox transformations.
- **Seasonal adjustment:**
  - Seasonal differencing.
  - Removing deterministic seasonal components.

Ergodicity: Practical View

- Directly testing ergodicity is difficult and rare in applied work.
- Common approach:
  - Assume a model (e.g. ARMA, GARCH).
  - Use known theoretical conditions for ergodicity of that model.
- If the model implies ergodicity, then:
  - Sample mean is a consistent estimator of the true mean.
  - Empirical ACF estimates the theoretical ACF.

It is still important to check for structural breaks, regime changes, and non-stationarities that may violate the assumptions.

Summary

- A time series is one realization of an underlying stochastic process.
- **Stationarity** is about invariance of distributions (or moments) under time shifts.
  - Strict vs weak stationarity.
  - Many models require at least weak stationarity.
- **Ergodicity** links time averages to ensemble averages.
  - Justifies using long-run sample averages as estimators of the true moments.
- In applications:
  - Always diagnose non-stationarity.
  - Use transformations or differencing if needed.
  - Choose models with theoretical guarantees of stationarity and ergodicity when possible.

Takeaways for Practice

- Never blindly assume stationarity; check data and context.
- Think about the mechanism: can it reasonably be stable over the sample?
- Use appropriate tests, but interpret them with care.
- When modelling, ensure that estimated parameters fall in the stationary region.
- Remember that all inference from one time series path implicitly relies on ergodic-type arguments.

Further Reading

- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control*. Wiley.
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications*. Springer.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Springer.
- Peters, O. (2019). "The Ergodicity Problem in Economics." *Nature Physics*, 15, 1216–1221.

Spectral View (Optional)

For a weakly stationary process with ACVF $\gamma(h)$, the spectral density is

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\lambda h}, \quad \lambda \in [-\pi, \pi].$$

- Encodes how variance is distributed over frequencies.
- For ARMA processes, $f(\lambda)$ has a closed-form expression in terms of the AR and MA polynomials.

Stationarity is required for the spectral representation to make sense.

Conditions for Stationary AR(1)

Consider

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.}(0, \sigma_\varepsilon^2).$$

Key facts:

- If $|\phi| < 1$, there exists a unique strictly stationary solution:

$$X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}.$$

- Then:

$$E[X_t] = 0, \quad \mathrm{Var}(X_t) = \frac{\sigma_\varepsilon^2}{1 - \phi^2}.$$

- If $|\phi| \geq 1$, variance is infinite or explodes, and no weakly stationary solution exists.

Thank You

Questions or discussion?