Why is stationarity so important?

Remember what stationarity means:

- **Covariance Stationarity (CS)**
  $\{y_t\}_{t=0}^{\infty}$ is called covariance stationary iff

$$\mathbb{E}[y_t] := \mu_t = \mu \text{ constant over time} \quad \text{and} \quad \text{Cov}(y_t, y_{t+k}) = \gamma(k) \text{ is a function of } k \text{ and NOT of } t \quad (\Longrightarrow \mathbb{V}(y_t) = \sigma^2 \text{ constant over time})$$

  (in words: constant expected value and variance; the autocovariance just depends on the time difference between two realizations and not on the time itself; this kind of stationarity does not assume anything about the underlying distribution, i. e. the distribution might change over time.)

- **Strictly Stationarity (SS)**
  $\{y_t\}_{t=0}^{\infty}$ is called strictly stationary iff

$$F_{Y_1,\dots,Y_k}(y_t, \dots, y_{t+k}) \equiv F_{Y_1,\dots,Y_k}(y_s, \dots, y_{s+k})$$

  (in words: every arbitrary joint distribution (of a finite number of random variables) is identical to the future joint distribution of the same number of random variables).

  This implies (if all moments exist) that all moments are constant and therefore strictly stationarity implies (if the first two moments exist) covariance stationarity.

Remember what we are interested in:

Although there is **only one realization** in each period, the aim is to find the underlying distribution in **each** period, the **ensemble distribution** (which we need for forecasts, statistical tests, cointegration, etc.). But how can we get from a sample with size one a whole distribution?

The theoretic (empirically unprovable) background is called ergodic theorem.

For illustration, assume there are two trajectories x_1 and x_2 each consisting of four elements. Every element was drawn from the constant (cf. picture) true, underlying (marginal) **ensemble density** $f_e(x)$ which we do not know, i. e. in particular, we do not know whether the realization is near the true mean μ_e or far away (the same with the true variance σ_e^2). Therefore, the basic assumption for the existence of such a constant distribution is a „constant" time series, i. e. a stationary time series.
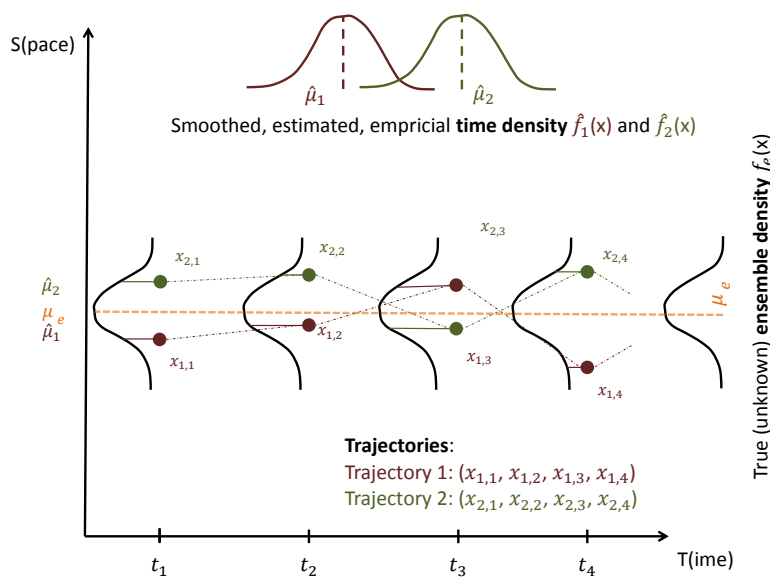
Although we have proven the existence, there is still the problem how to estimate it. Since we have more than one realization in time, we can build the empirical **time densities** $\hat{f}_1(x)$ and $\hat{f}_2(x)$ with average values $\hat{\mu}_1$ and $\hat{\mu}_2$ and empirical variances $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$. The connection between the moments of the **given, empirical time densities** and the **unknown, underlying ensemble density** is made by the **Ergodic Theorem**:

Assumptions: For an asymptotically covariance stationary time series, e. g. x_1, which is ergodic (not empirically provable), it holds that

$$\lim T \to \infty \, \mathbb{E}\left[\left(\underbrace{\frac{1}{T}\sum_{t=1}^{T} x_{1t}}_{\hat{\mu}_1} - \mu_e\right)^2\right] = 0 \quad \text{where „ergodicity for the mean" means:} \quad \frac{1}{T}\sum_{t=1}^{T} x_{1t} \xrightarrow{P} \mu_1 \text{ exists}$$

$$\lim T \to \infty \, \mathbb{E}\left[\left(\underbrace{\frac{1}{T}\sum_{t=1}^{T} (x_{1t} - \hat{\mu}_1)^2}_{\hat{\sigma}_1^2} - \sigma_e^2\right)^2\right] = 0 \quad \text{where „ergodicity for the variance" means:} \quad \frac{1}{T}\sum_{t=1}^{T} (x_{1t} - \hat{\mu}_1)^2 \xrightarrow{P} \sigma_1^2 \text{ exists}$$

which means that for increasing sample size T the time average value $\hat{\mu}_1$ converges to the ensemble mean μ_e (the same for variances $\hat{\sigma}_1^2$ and σ_e^2) and therefore connects the moments of those a priori different densities. An important sufficient condition for both ergodicity for the mean and variance for Gaussian processes is $\sum_{j=0}^{\infty} |\gamma(j)| < \infty$ (in this case, this condition even ensures ergodicity for all moments).

Finally, remember that the basic assumption is a form of **stationarity**.