Stationarity

Technical Definition

Strong Stationarity is a property of a random process that the joint probability distribution does not change with time. This is a very strong assumption. A relaxed version of this assumption is of the weak stationarity. Now rather that talking about the entire distribution, we constrict ourselves to the first two moments of a multivariate distribution. The first two moments are the vector of means and the co-variance matrix. Therefore, weak stationarity assumption is that:

1. The mean $\mu_t = \mu$ does not change with time.
2. The variance $\sigma_t^2 = \sigma^2$ does not change with time.
3. The autocovariance function $\gamma(t, t-k) = \gamma(k)$ only depends on the time difference and does not change with time.

Intuitive Definition

Stationarity means that statistical properties of a process do not change over time. Let us understand this step by step.

- A realisation of a time series (what we observe) falls out of a certain probability distribution which is usually unknown.
- The sequence of such realisations is called a stochastic process. Since we have a sequence of realisations, we need to talk about joint probability distribution of the realisations at different time instants.
- However, these realisations might not be independent of each other. For example, my mood right now is surely dependent on the mood I had one hour back. Moreover, the joint probability distribution could be a function of time itself. Overtime, I am learning to take things less seriously and more sincerely. Hence, my mood is becoming more and more stable.
- A fundamental part of predicting something is to understand the underlying probability distribution. If the probability distribution is changing with time, it is very difficult to predict the future.
- Therefore, we need to have that the joint probability distribution is not a function of time. This is the stationarity assumption.
- This assumption is very strong and is rarely satisfied in practice. We silly humans need to hammer this randomness down to fit the linear equations which are easier to work with. Therefore, we need to relax this assumption to a weaker version of stationarity and only restrict ourselves to the first two moments of the joint probability distribution.

Stationarity through an example

Let us talk about an AR(1) process which is the simplest class of linear stochastic processes. Imagine a drunkard walking. This drunkard has been hammered with some good Russian vodka. Now, the goal is to predict where this drunkard will land after taking certain number of steps, say 100 steps. We have a structure here! His each step is dependent on his previous step plus some random noise. The drunkard can only stretch his legs within some bounds and given the impact of comrade standard vodka, the next movement of his feet is independent of his previous movement. Hence, we can safely assume that this random noise is white noise that is it has no structure or is random, atleast in a linear sense. Formally, we can write this as:

$$X_{t+1} = X_t + e_{t+1}$$

X_t is the location of the drunkard at time t and e_{t+1} is the white noise with zero mean, σ^2 variance and zero covariance between succesive positions.

Each step of the drunkard adds some randomness to his trajectory. The marginal increase in his distance has a variance of σ^2. This variance keeps on cascading with each succesive step. The joint probability distribution here is a function of time.

$$\begin{aligned}
\mathrm{Var}(Y_t) &= \mathrm{Var}(e_1 + e_2 + \ldots + e_t) \\
&= \mathrm{Var}(e_1) + \mathrm{Var}(e_2) + \ldots + \mathrm{Var}(e_t) \text{ (independence)} \\
&= \sigma^2 + \sigma^2 + \ldots + \sigma^2 = t\sigma^2
\end{aligned}$$

Let us look at a simulatuion of this process. We will first define a function that will simulate the drunkard's walk. By the way, this is called a random walk. Maybe, the person who coined the term thought of the same example.

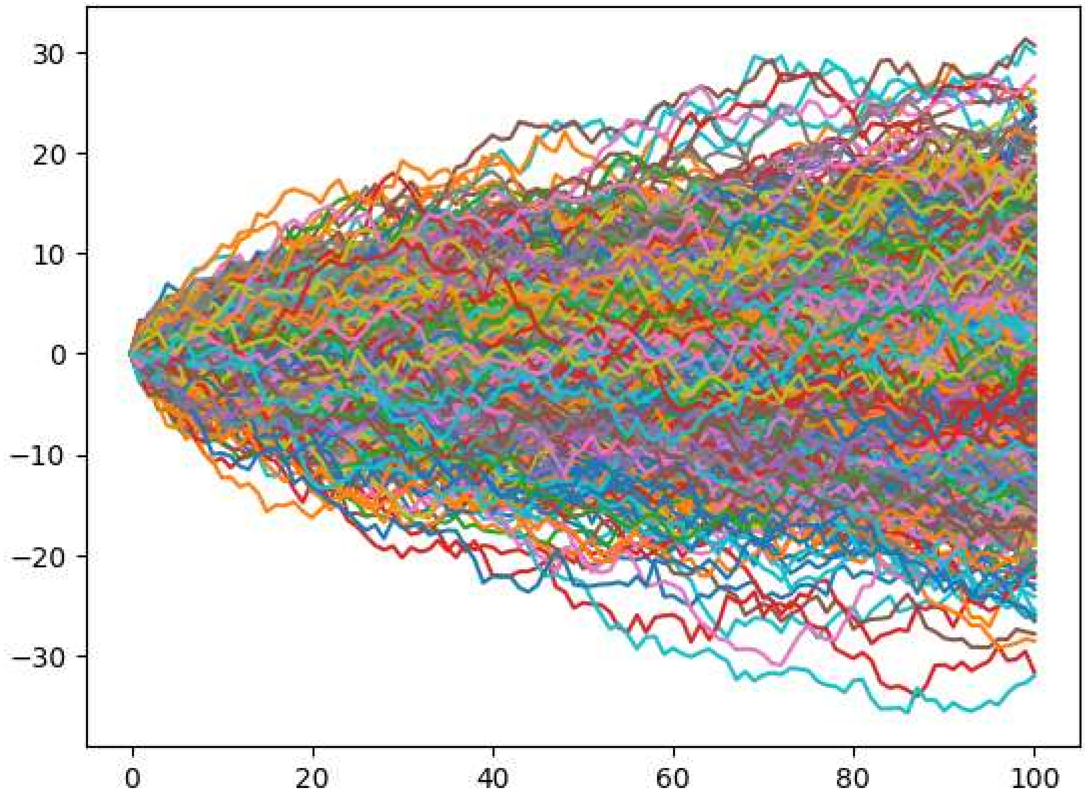
In [ ]:

```
import numpy as np
# Simulate a randomwalk
def generate_random_walk (steps: int, initial=0) -> list:
  walk = np.zeros(steps+1)
  walk[0] = initial
  for i in range(1, steps+1):
    walk[i] = walk[i-1] + np.random.normal(0, 1)
  return walk
```

Let us simulate thousand of these drunkards, each taking 100 steps. You can see the plot below and appreeciate how the variance increases with the number of steps.

In [ ]:

```
# Plot these 1000 random walks
import matplotlib.pyplot as plt
for _ in range(1000):
  plt.plot(generate_random_walk(100))
```



There are formal and not so formal ways to check for stationarity. The not-so-formal way is to just look at the time series plot and see if mean and variance is changing with time. The formal way is to do a statistical test. The most popular statistical test is the Augmented Dickey-Fuller test. The null hypothesis of this test is that the time series is non-stationary. If the p-value is less than 0.05, we reject the null hypothesis and conclude that the time series is stationary.

Dealing with non-stationarity

Random walk (also called unit-root processes) are a type of non-stationary process. Many wise women and men have claimed that the movement of many stock prices follow a random walk. For a random walk, the difference of successive realisations is a white noise which is stationary. We just discovered a way to make a process stationary. Differencing a time series is one of the most popular ways to make a non-stationary process stationary. The intuition is that if the process is non-stationary, the difference of successive realisations will be stationary. The transformation of stock prices into returns is an example of proportional differencing. The returns series is usually stationary, unless you are unlucky or are looking crypto-currencies.