The Analysis of Time Series Data as a Signal

Stochastic Processes

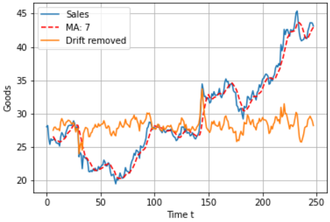
IEOR 4574: Forecasting: *A Real-World Application*

Syed Waseem Haider, Ph.D.

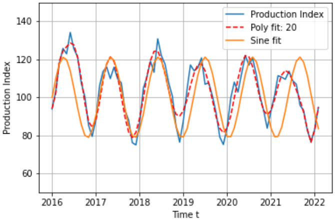
Dr.Waseem.Haider@gmail.com

January 23, 2023

Motivation behind Discussing the Science of Time Series



(a) A retailer Sales time series data



(b) Ice cream Sales distribution data

- What's wrong with the two time series?
- How to characterize the time series?
- How to fix the two time series?

Forecasting Introduction

Agenda for learning

- The material covered here intends to familiarize with the mathematical concepts at the heart of time series data.
- It will answer the questions such as:
  - What is a time series?
  - What are the probabilistic properties of a time series that help in solving them to create value in real life?
  - What are the different types of behaviors or dynamics that exist in a time series?
  - Is there another way to observe time series other than time?
  - How do time series occur in real-life situations?
- Recommended textbook: Introduction to Probability Models by Sheldon Ross. *Any version will do.*

Forecasting Introduction

Key concepts

- Random Event.
- Random Variable.
- Mean, Variance, Autocorrelation, and Autocovariance.
- Random or Stochastic Process.
- Stationary Random Process.
- Conditional and Marginal Probabilities.

Time Series

Introduction

- Time series y_t, $t = 1, \ldots, n$, could be the sales, or shipments of a product on daily, weekly or monthly cadence.
- The time series data is split into the training and test data.
- For training the model:
  - The model is fitted or trained on $t = 1, \ldots, k-1$.
  - $\hat{y}_t$ is a fitted time series (training) output of a model such as SARIMAX, GBRT, Facebook Prophet, etc.
  - The residual error at time t is $e_t = y_t - \hat{y}_t$.
- For testing the fitted model:
  - $\hat{y}(t)$ is the model predictions or forecasts on the test data, $t = k, \ldots, n$.
  - The prediction (forecast) errors are $e(t) = y_t - \hat{y}(t)$.
- Usually, the residuals are smaller than the forecasted errors. These definitions allow distinguishing residuals in the training data from the prediction errors in the test data, or actuals into the future.

Time Series

Introduction

- A time series is usually the occurrence of random events in the order in time.
- These events can be discrete such as the arrival of customers at a bank for servicing their accounts.
- Events can also be continuous over time such as the physical activity recorded by the gyroscope sensors in the smartwatch for the patient in a clinical trial.
- A **Random Variable** (RV) represents the random events observed from a sample space (discrete/continuous, and finite/infinite), e.g., rolling of a dice, or stock prices. It is denoted as X.
- There is a probability distribution associated with the RV X, such that $P[X = x]$ with the parameters of the distribution.
- The mean of a random variable $X \epsilon \{0, \ldots, N\}$ is defined as $E[X] = \sum^{N} i \cdot P[X = i]$.

- The variance of a random variable $X \epsilon \{0, \ldots, N\}$ is defined as
$$VAR[X] = E[(X - E[X])^2] = \sum_{i=0}^{N} (i - E[X])^2 \cdot P[X = i].$$
- For example, $X \sim \mathcal{N}[\mu, \sigma^2]$ suggests that the random variable X is Normally (Gaussian) distributed with parameters mean $E[X] = \mu$, and variance $VAR[X] = \sigma^2$.
- **Random process** is a Random Variable (RV) observed over time, e.g., rolling of a dice, or stock prices observed over time. Denoted as $X(t)$.
- **Stochastic process** is another common name for the random process.
- The probability that at time t there will be x customers in the bank is $P[X(t) = x]$, or simply $P[X_t = x]$.
- The time series can be formally described as $\{X_t\}$, where $X_t \epsilon \{H, T\}$, or $X_t \epsilon \mathbb{R}$.

Time Series as a Stochastic Process X(t)

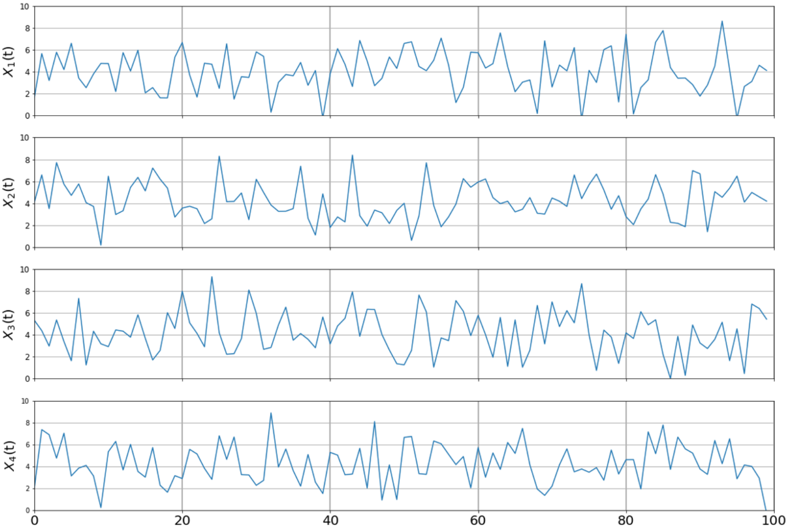
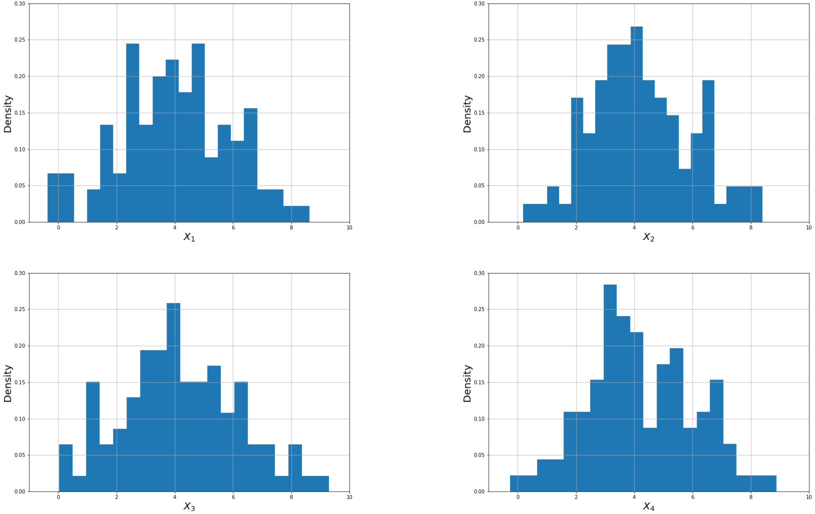


Figure: A Random Process $X(t) = A$ where $A \sim \mathcal{N}[4, 2]$.

Distributions from the Stochastic Processes



- Bernoulli is a common discrete random variable. The RV $X \epsilon \{0, 1\}$ is defined over two events such that $P[X = 1] = p$, and $P[X = 0] = 1 - p$.
  - The mean of X, $E[X] = p \cdot 1 + (1 - p) \cdot 0 = p$.
  - And the variancef X, $Var[X] = p \cdot (1 - p)^2 + (1 - p) \cdot (0 - p)^2 = p(1 - p)^2 + (1 - p)p^2 = (1 - p)p(1 - p + p) = (1 - p)p$.
  - p is the only parameter of the distribution.

Some Common Distributions

Binomial Distribution

- The Binomial random variable $X \epsilon \{0, \ldots, n\}$ over n independent random trials suggests i successes with the probability $P[X = i] = \binom{n}{i}\ p^i\ (1 - p)^{n-i}$, where p is the probability of success.
- The Binomial random variable is a discrete random variable.
  - The mean of X, $E[X] = np$.
  - The Variance of X, $Var[X] = np(1 - p)$.
  - n and p are the parameters of the distribution.
- The bagging techniques are based on the Binomial distribution, where data is randomly drawn with replacement from the data set to build the models such as **Random Forest** regression.
- The replacement allows for each draw to be independent of the other draws.

Some Common Distributions

Geometric Distribution

- The Geometric is an interesting distribution. It describes a discrete random variable $X \epsilon \mathbb{N}$ over n independent random trials such that at the $n{-}th$ trial occurs the first success with the probability $P[X = n] = (1-p)^{n-1}p$, where p is the probability of success.
- In other words, a sequence of $n-1$ continuous failures precedes success.
  - The mean of X, $E[X] = \frac{1-p}{p}$.
  - The Variance of X, $Var[X] = \frac{1-p}{p^2}$.
  - p is the only parameter of the distribution.
- Examples of Geometric distribution include, how many components to process before arriving at the first failure. How many shipments are successfully delivered before a breakdown in the truck happens?

Some Common Distributions

Poisson Distribution

- The Poisson distribution defines a discrete random variable $X \epsilon \mathbb{Z}$ such that there are i events with the probability $P[X = i] = e^{-\lambda} \frac{\lambda^i}{i!}$, where λ is the rate at which events are occurring.
  - The mean of X, $E[X] = \lambda$.
  - The Variance of X, $Var[X] = \lambda$.
  - λ is the only parameter of the distribution.
- Examples of Poisson distribution include the number of errors on a single page. The number of missed shipments on a truck route.

- A new programmer commits programming errors at the rate of 10 syntax errors each hour that she codes. If the programmer has coded for 40 hours, what is the probability that she will make 410 errors?
- The programming error rate is $\lambda = 10 \times 40 = 400$.
- $P[X = 410] = e^{-\lambda}\frac{\lambda^{410}}{410!} = 0.017$

Some Common Distributions

Normal (Gaussian) Distribution

- The Normal distribution defines a continuous random variable $X \epsilon \mathbb{R}$ such that the probability that an event $X \leq b$ will happen is

  $$P[X \leq b] = \int_{-\infty}^{b} f_X(x)\ dx.$$

- $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ is the probability density function (PDF) of the normal distribution.
  - The mean of X, $E[X] = \int_{-\infty}^{\infty} x\ f_X(x)\ dx = \mu$.
  - The Variance of X, $Var[X] = \int_{-\infty}^{\infty} (x-\mu)^2\ f_X(x)\ dx = \sigma^2$.
  - μ and σ are the two parameters of the distribution.
- The PDF of Normal distribution is the famous bell curve.
- The PDF is symmetric around the mean, μ, and the width of the bell is given by σ the standard deviation.

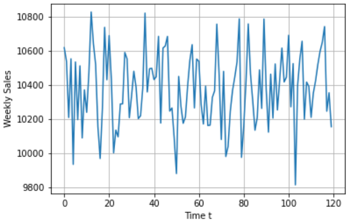
- Let $Y = X + \alpha$ be a sensor system. Y is the sensor observations, $X \sim \mathcal{N}[0, 2\sqrt{2}]$ is the latent signal, and α is the bias in the sensor. Find the value of α such that $P[Y \leq 1] = 0.5$?
- $P[Y \leq 1] = P[X + \alpha \leq 1] = P[X \leq 1 - \alpha] = 0.5$.
- $P[X \leq 1 - \alpha] = \int\limits_{-\infty}^{1-\alpha} f_X(x)\ dx = 0.5$.
- $f_X(x) = \frac{1}{\sqrt{2\pi(2\sqrt{2})^2}} e^{\frac{-(x-0)^2}{2(2\sqrt{2})^2}} = \frac{1}{4\sqrt{\pi}} e^{\frac{-1}{16}x^2}$.
- One can solve the integration in $P[X \leq 1 - \alpha]$ to find α by substituting the PDF in the equation.
- However, there is another intuitive, easier way to find $P[X \leq 1 - \alpha]$.
- We know up to the mean $\mu = 0$, $P[-\infty \leq X \leq 0] = 0.5$. Or simply $P[X \leq 0] = 0.5$.
- Hence, for $P[X \leq 1 - \alpha] = 0.5 \Rightarrow$ bias, $\alpha = 1$

Stationarity in Time Series
Stochastic Process

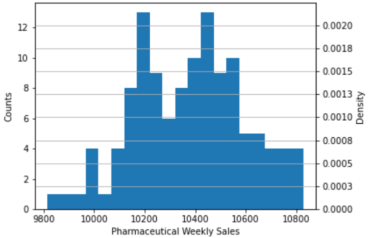
- **Stationarity** in a random process implies that the probability of a given sequence of observations is independent of time itself, but will depend on the distances in time among those observations. It's expressed as a joint Cumulative Density Function (CDF) such that $F_{X_{t_1},\ldots,X_{t_n}}(x_1,\ldots,x_n) = F_{X_{t_1+\tau},\ldots,X_{t_n+\tau}}(x_1,\ldots,x_n)$.
- **Stationarity** also requires that the first-order CDF is independent of time, $F_X(x;t) = F_X(x)$. Subsequently, the mean, $E[X(t)] = \mu_X$, and the variance, $VAR[X(t)] = \sigma^2{}_X$, are also independent of time.

- **Stationarity** makes the second-order CDF a function of time interval than the time itself, $F_{X_{t_1},X_{t_2}}(x_1, x_2) = F_{X_{t_1+\tau},X_{t_2+\tau}}(x_1, x_2)$.
- The autocorrelation, $E[X_{t_1}X_{t_2}] = R_X(t_1, t_2) = R_X(t_2 - t_1) = R_X(\tau)$, and autocovariance $E[(X_{t_1} - \mu_X)(X_{t_2} - \mu_X)] = C_X(t_1, t_2) = C_X(t_2 - t_1) = C_X(\tau)$ are also functions of time interval, τ.
- The above conditions make the stochastic process a **Strict-Sense Stationary process**. A more relaxed **Wide-Sense Stationary process** will also be discussed.

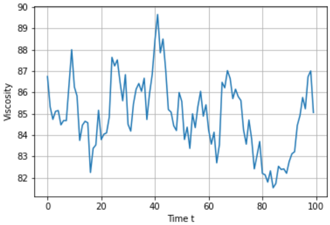
Stationary Time Series



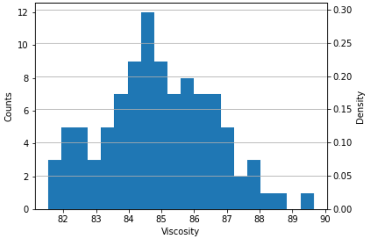
(a) Sales time series



(b) Sales distribution

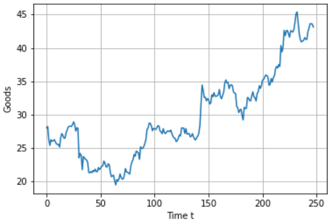


(c) Viscosity time series

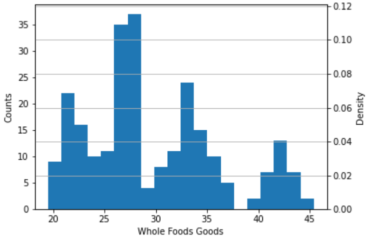


(d) Viscosity distribution

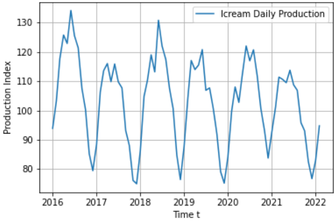
Non-Stationary Time Series



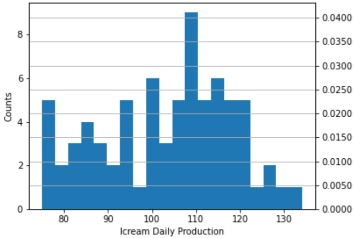
(a) Goods time series



(b) Goods distribution



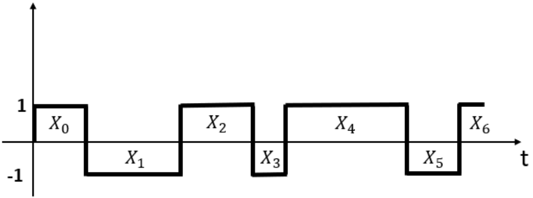
(c) Ice Cream Frozen Desert time series



(d) Ice Cream Frozen Desert distribution

Time Series from a Poisson Process

- **A random process** $X_t \in \{-1, 1\}$ at any t. The probability of $X_0 = 1$, or $X_0 = -1$ is $P[X_0 = \pm 1] = \frac{1}{2}$.
- $X(t)$ alternates between -1 and 1 with every event that occurs with a Poisson process at rate λ. For example, if $X_t = 1$, it will become $X_t = -1$ when the next event occurs following $\sim \mathcal{P}[\lambda]$.



- Question: If $X_0 = -1$, what is the probability that $X_t = -1$?.

Time Series from Poisson Process

A detour into the probability theory

- Before answering the question above, one must understand conditional probability principles.
- Probability of an event A is P(A). Now, knowing that the event B has happened, the probability of event A is called the conditional probability, $P(A|B)$.
- $P(A|B) > P(A)$ if happening of B makes happening of A more likely.
- Similarly, $P(A|B) < P(A)$ if happening of B makes A less likely.
- So, the information about A changes with information about B if the two have a relationship.
- For example, the knowledge of whether it is cloudy or sunshine outside, changes one's guess if it will rain or not than if one did not know the outside conditions.

Time Series from Poisson Process

A detour into the probability theory

- $P(A|B) = \frac{P(A,B)}{P(B)} \Rightarrow P(A,B) = P(A|B)P(B)$.
- Conditional probabilities also allow a framework to model cause and effect relationships of the events (the Bayesian paradigm), or the ordering in time among the events (the Markovian methods).
- Marginal probabilities are when two RVs X and Y are observed together with the situation that if X could be averaged out then the marginal probability of Y, $P[Y] = \sum_x P[Y,X] = \sum_x P[Y|X]P[X]$.

Time Series from Poisson Process

Coming back to answer the question

- Restating the Question again: If $X_0 = -1$, what is the probability that $X_t = -1$?
- Answer: $P[X_t = -1|X_0 = -1]$.
- Similarly, $P[X_t = 1|X_0 = 1]$, $P[X_t = 1|X_0 = -1]$, and $P[X_t = -1|X_0 = 1]$.
- The probability of X_t is described as $P[X_t = 1] = P[X_t = 1|X_0 = -1]P[X_0 = -1] + P[X_t = 1|X_0 = 1]P[X_0 = 1]$.
- The probability of X_t is described as $P[X_t = -1] = P[X_t = -1|X_0 = -1]P[X_0 = -1] + P[X_t = -1|X_0 = 1]P[X_0 = 1]$.
- In summary, the probability of X_t is described as
$$P[X_t = \pm 1] = \sum_{i=-1}^{1} P[X_t = \pm 1|X_0 = i]P[X_0 = i].$$
- When $X_t = X_0$ the probability is defined as $P[X_t = X_0]$, or it is simply represented as $P[X_t = \pm 1|X_0 = \pm 1]$.

Time Series from Poisson Process

- Similarly, $X_t \neq X_0$ the probability is defined as $P[X_t \neq X_0]$, or it is represented as $P[X_t = \pm 1 | X_0 = \mp 1]$.
- Assuming $X_0 \pm 1$, the probability of the Poisson process X_t depends on if there has occurred an even or odd number of events since $t = 0$.
  - Even number of events: $P[X_t = \pm 1 | X_0 = \pm 1] = \sum_{j=0}^{\infty} \frac{(\lambda t)^{2j}}{(2j)!} e^{-\lambda t} =$ $e^{-\lambda t} \ \frac{1}{2} \left(e^{\lambda t} + e^{-\lambda t} \right) = \frac{1}{2} \left(1 + e^{-2\lambda t} \right)$
  - Odd number of events: $P[X_t = \pm 1 | X_0 = \mp 1] = \sum_{j=0}^{\infty} \frac{(\lambda t)^{2j+1}}{(2j+1)!} e^{-\lambda t} =$ $e^{-\lambda t} \ \frac{1}{2} \left(e^{\lambda t} - e^{-\lambda t} \right) = \frac{1}{2} \left(1 - e^{-2\lambda t} \right)$
- Using the two Taylor series expansions $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$, and $e^{-x} = 1 - \frac{x}{1!} + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots$.
- When the two series are added together, it gives the even terms only. Similarly, when the two are subtracted result in odd terms only.

Time Series from Poisson Process

- For example, consider the case of $\left(e^{\lambda t} + e^{-\lambda t}\right)$:
  - $e^x + e^{-x} = \left(1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots\right) + \left(1 - \frac{x}{1!} + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots\right)$
  - $\Rightarrow e^x + e^{-x} = 2 + \frac{2x^2}{2!} + \frac{2x^4}{4!} + \dots = 2\left(1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots\right)$
  - $\frac{1}{2}\left(e^x + e^{-x}\right) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots = \sum_{j=0}^{\infty} \frac{(\lambda t)^{2j}}{(2j)!}$
  - Similarly, $\left(e^{\lambda t} - e^{-\lambda t}\right)$ is solved.
- $P[X_t = 1] =$
  $P[X_t = 1 | X_0 = -1] P[X_0 = -1] + P[X_t = 1 | X_0 = 1] P[X_0 = 1]$
- $P[X_t = 1] = \frac{1}{2}\left(1 - e^{-2\lambda t}\right)\frac{1}{2} + \frac{1}{2}\left(1 + e^{-2\lambda t}\right)\frac{1}{2} = \frac{1}{2}$
- $P[X_t = -1] = 1 - P[X_t = 1] = \frac{1}{2}$

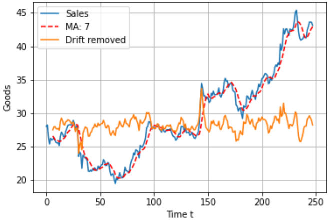
Time Series from Poisson Process

- **Mean** of X_t:
  $E[X_t] = (-1)P[X_t = -1] + (1)P[X_t = 1] = (-1)\frac{1}{2} + (1)\frac{1}{2} = 0.$
- **Variance** of X_t:
  $E[(X_t - E[X_t])^2] = E[(X_t)^2] = (-1)^2 P[X_t = -1] + (1)^2 P[X_t = 1] = 1.$
- Because it's a zero-mean process, $E[X_t] = 0$, the **autocovariance** is the same as **autocorrelation**.
  $COV(X_t, X_{t+\tau}) = E[(X_t - E[X_t])(X_{t+\tau} - E[X_{t+\tau}])] =$
  $E[(X_t - 0)(X_{t+\tau} - 0)] = E[(X_t)(X_{t+\tau})] = R_X(\tau).$
- The autocorrelation value is computed for two cases viz., $X_{t+\tau} = X_t$, then the product of the two will always be '1'. For example, if $X_t = -1$ and $X_{t+\tau} = -1$, then $(X_t)(X_{t+\tau}) = 1$.
- Similarly, when $X_{t+\tau} \neq X_t$, then the product will be always be '-1'.

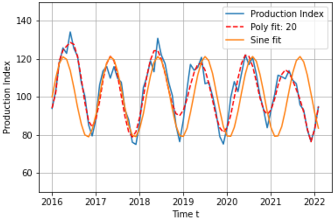
Time Series from Poisson Process

- Hence, $R_X(\tau) = E[(X_t)(X_{t+\tau})] = (1)\ P[X_t = X_{t+\tau}] + (-1)\ P[X_t \neq X_{t+\tau}] = -\frac{1}{2}\left(1 + e^{-2\lambda\tau}\right) - \frac{1}{2}\left(1 - e^{-2\lambda\tau}\right) = e^{-2\lambda\tau}$
- This Poisson process satisfies all the criteria of a stationary stochastic process.
- It is also a continuous-time Markov chain process because it only depends on the present to determine future events.

Motivation behind Discussing the Science of Time Series



(a) A retailer Sales time series data



(b) Ice cream Sales distribution data

- What's wrong with the two time-series? Some elements are causing non-stationarities in the time series.
- How to characterize the time series? A random process with some kind of underlying probability distribution driving the behavior over time.
- How to fix the two time-series?

Sources of Non-Stationarities in Time Series

Methods to make time-series stationary

- **Drift** Is the result of some kind of bias in the data that shifts the time series data over time.
- Examples of drifts include a bias in the sensor, the direction of the wind changing the course of an airplane, and a selling frenzy of the stocks on a particular day.
- **Drift** in Time series can be removed by fitting a Moving Average (MA) of lag n, and then subtracting the fitted MA from the actual time series.
- **Drift** in Time series can also be removed by differencing such as Autoregression of lag 1. For example, $y_t - y_{t-1} = y_{t-1} - y_{t-2} + b - b + \epsilon_t - \epsilon_{t-1}$, a popular technique used in ECG signals.

Sources of Non-Stationarities in Time Series

Methods to make time-series stationary

- **Trend** is an unanticipated behavior that could lead to linear or nonlinear growth or decay in the time series data over time. In business, it's a term that covers noticeable, but sometimes difficult to attribute, exogenous behaviors.
- Changes in travel behavior over time due to COVID impact, unattributable uncertainty in the market impacts stock prices, etc.
- **Trend** in Time series can be removed by fitting a first-degree polynomial (a linear model) and removing the slope m. For example, $y_t - x_t = mt - \hat{m}t + b + \epsilon_t$, a popular technique used in ECG signals. For example, a bad resistor in a device can cause a deterministic trend when measuring voltage.

- **Seasonality** is the repeated cycles of one or more behaviors present in the time series data.
- Cycles are usually characterized by amplitude and frequency and mathematically represented by sinusoidal functions.
- Examples of seasonality are seasonal flu, shopping patterns on Thanksgiving, consumption of ice cream, and holiday travel in the Summer.
- **Seasonality** in Time series can be removed by first fitting either a high degree polynomial or a sine function, and then removing that from the time series, which gives the delta sales, or shipments higher or less than the seasonal behavior.

- Fitting a regression model $z_t = \beta_0 + \beta_1 sin(\omega t) + \beta_2 cos(\omega t)$, which is actually a sine wave with the phase α, $z_t = \beta sin(\omega t + \alpha)$, $\Rightarrow z_t = \beta cos(\alpha) sin(\omega t) + \beta sin(\alpha) cos(\omega t)$, where $\beta_1 = cos(\alpha)$, and $\beta_2 = sin(\alpha)$.
- For example, the sales that is cleansed seasonal impact is $Y_t = ConsumerDemand_t + Distribution_t + Price_t + BrandLoyalty_t - z_t$, where z_t is the fitted model to the seasonal sales.
- A Sine wave stochastic process is non-stationary. For example, if $\beta \sim \mathcal{U}[0, 1]$ in $z(t) = \beta\ sin(\omega t)$ then the probability density function, is also a function of time $f_Z(z, t) = \frac{1}{sin(\omega t)} f_\beta \left(\frac{z}{sin(\omega t)} \right)$.

Non-stationarity in Time Series

Seasonality

- Sine and Cosine harmonics of the fundamental frequency can be added to fit a complex seasonality such as both annual and monthly seasonality.

- To model an annual seasonality using monthly data, the time period is $T = 12$, and the frequency is $f = \frac{1}{T}$. The annual seasonal model is $z_t = \beta_0 + \beta_1 sin(\frac{2\pi}{12}t) + \beta_2 cos(\frac{2\pi}{12}t)$.

- It can be extended to model both annual and half-early as $z_t = \beta_0 + \beta_1 sin(\frac{2\pi}{12}t) + \beta_2 cos(\frac{2\pi}{12}t) + \beta_3 sin(\frac{2\pi}{6}t) + \beta_4 cos(\frac{2\pi}{6}t)$

- Similarly, the quarterly seasonal model using the monthly data is $z_t = \beta_0 + \beta_1 sin(\frac{2\pi}{3}t) + \beta_2 cos(\frac{2\pi}{3}t)$.

Fixing Non-Stationarity by Removing Drift

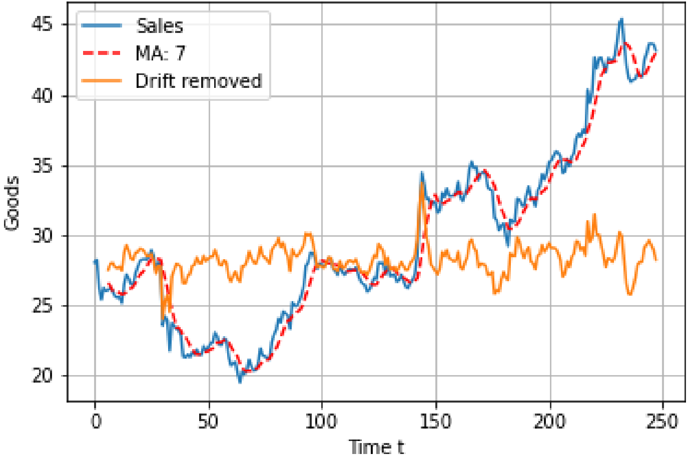
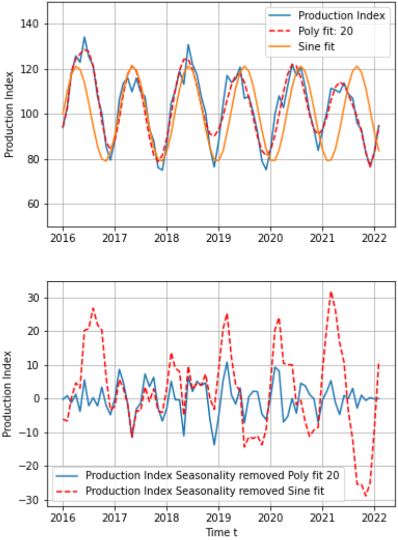


Figure: Goods time series

Fixing Non-Stationarity by Removing Seasonality



Wide-Sense Stationary Process and Autocorrelation

- **Wide-Sense Stationary (WSS) Process** is a loosely defined stationary process in situations where establishing stationarity is not easy; it suffices if the mean is independent of time, $E[X(t)] = m$, and the autocorrelation (also autocovariance) function depends only on the time difference instead of a particular time, $R_X(\tau)$ (by extension autocovariance is also only dependent on the time difference).
- There are several properties of $R_X(\tau)$ of a WSS that helps the analysis of the data.
- First, $R_X(\tau)$ when $\tau = 0$, $R_X(0) = E[X(t)^2]$ provides the average power in the signal (e.g. $E[P(t)] = E[I(t)^2]R = \frac{E[V(t)^2]}{R}$.) The second moment is usually associated with the average power.
- Second, For real-valued random process, $R_X(\tau)$ is an even function, i.e. $R_X(\tau) = R_X(-\tau)$.

Wide-Sense Stationary Process and Autocorrelation

- Third, $R_X(\tau)$, is a measure of the average rate of change of a random process. If the signal changes slowly over time, it remains correlated with itself over a longer period, whereas if it changes quickly, it is correlated with itself for a shorter duration.
- Fourth, $R_X(\tau)$ is maximum at $\tau = 0$.
  - Following the Cauchy–Schwarz inequality in Euclidean space, $<u, v>^2 = <u, u> <v, v> \Rightarrow |<u, v>| = \sqrt{<u, u>} \sqrt{<v, v>} = ||u|| \; ||v||$, where $< \, , >$ is the inner product.
  - The idea behind Cauchy–Schwarz inequality is that if two non-zero signals are orthogonal then their dot product will be zero, but the product of their magnitudes will be greater than zero.
  - Both sides in Cauchy–Schwarz inequality will be equal if u and v are linearly independent. Linearly independence requires that $\alpha u + \beta v = 0$ only when $\alpha = 0$, and $\beta = 0$. In that case, $u = \gamma v$, and $|<\gamma v, v>| = \sqrt{<\gamma v, \gamma v>} \sqrt{<v, v>}$.

Wide-Sense Stationary Process and Autocorrelation

- $\left(\sum_{j=0}^{\infty}(x_j y_j)^2\right) \leq \left(\sum_{j=0}^{\infty} x_j^2\right)\left(\sum_{j=0}^{\infty} y_j^2\right)$,
  $\left(\sum_{t=0}^{\infty}(X(t)X(t+\tau))^2\right) \leq \left(\sum_{t=0}^{\infty} X^2(t)\right)\left(\sum_{t=0}^{\infty} X^2(t+\tau)\right) \approx$
  $E[X(t)X(t+\tau)]^2 \leq E[X^2(t)]E[X^2(t+\tau)] \Rightarrow R_X^2(\tau) \leq R_X^2(0)$.

- Fifth, for signals with seasonalities or repeated patterns, $R_X(\tau)$ is periodic, $R_X(\tau+d) = R_X(\tau)$ such as d is monthly, quarterly, half-yearly, annually, or more than one repeater pattern could be present. For example, the autocorrelation of the Production Index of ice cream can be approximated by $R_X(\tau) = cos2\pi ft = cos\frac{2\pi}{12}t$.
- Sixth, $R_X(\tau)$ can contain three types of components: i) a periodic component, ii) a component due to non-zero mean of the signal, and iii) a component that approaches zero as $\tau \to \infty$.

Autocorrelation

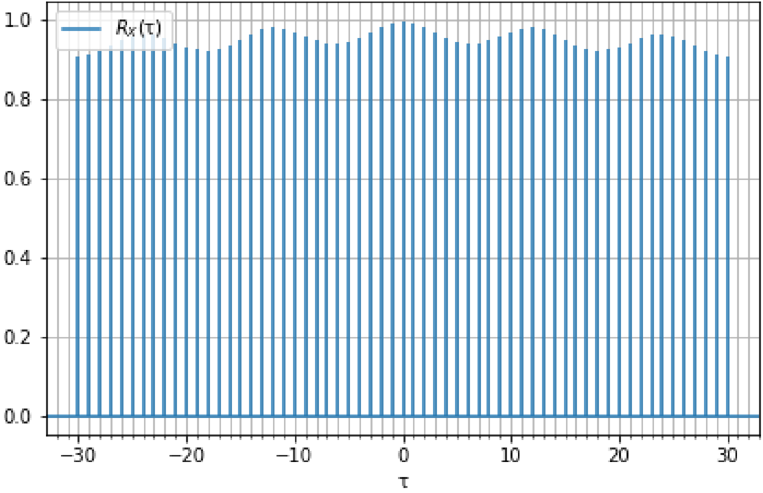


Figure: Autocorrelation of the Production Index of ice cream. Note period T=12 months, and the average power $R_X(0)$.

Power Spectral Density

- The Power Spectral Density (**PSD**) is the Fourier Transform of the autocorrelation function $R_X(\tau)$,
$S_X(f) = \mathcal{F}\{R_X(\tau)\} = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi ft} d\tau.$
- When $R_X(\tau)$ is real-valued, even function, $S_X(f)$ is also real-valued, even function of f.
- $\mathbf{h}^T S_X(f)\mathbf{h} \geq 0$ for all f. It can be proven using Eigenvectors and Diagonalization that $S_X(f)$ is a positive-semi definite matrix.
- PSD, $S_X(f)$, also helps to identify seasonalities or repeated behaviors present in the signal at different time lengths (periodicity/frequency) such as quarterly, half-yearly, annually, or existing one or more together.
- $R_X(\tau)$ when $\tau = 0$, $R_X(0) = E[X(t)^2]$ provides the average power in the signal over all frequencies,
$R_X(0) = \mathcal{F}^{-1}\{S_X(f)\} = \int_{-\infty}^{\infty} S_X(f) e^{-j2\pi f\tau} df = \int_{-\infty}^{\infty} S_X(f) df.$

Power Spectral Density

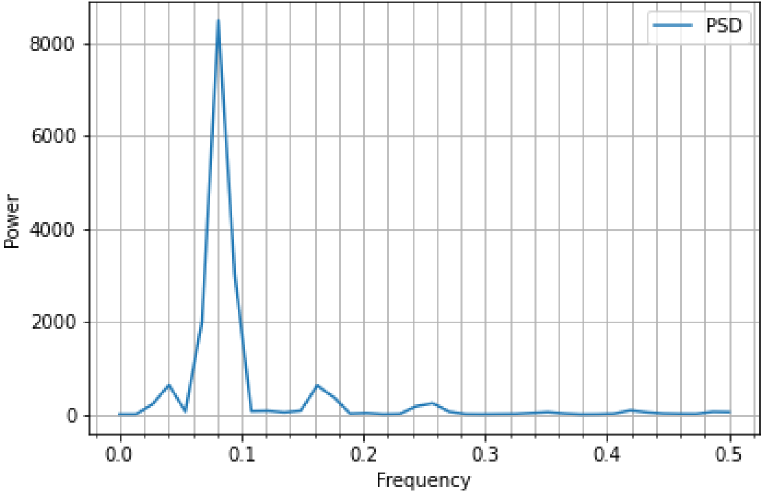


Figure: Power Spectral Density (PSD) of the Production Index of ice cream

Time Series as a Random Walk

- A **Random Walk** is the position of the player's winnings, $S(t) = \sum_{j=0}^{t-1} X(j) + X(t)$, when he plays a game following a random process $X(t)$. Up to the time $t-1$ the total winnings are $\sum_{j=0}^{t-1} X(j)$. At turn t, he can win $X(t) = \$+1$ or lose $\$-1$ with the probability p and $q = 1 - p$, respectively.
- When $p = q$, the random walk is called symmetric.
- Examples of the random walk are the prices of stocks on Monday mornings, the path of the molecule in gas or liquid (Brownian motion), the position of a gambler, and the path of an animal searching for food called foraging.

A Random Walk Example with Bernoulli Process

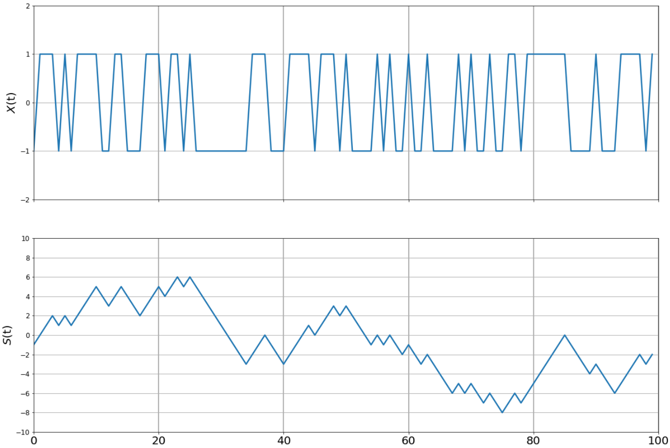


Figure: A random walk with $p = \frac{1}{2}$, $E[S(t)] = (2p - 1) \times t\ steps$.

Time Series as a Random Walk

- A simple random walk $S(t)$ naturally displays the Markovian property, that the current position $S(t-1)$ is sufficient for the next position $S(t)$, where $S(0), S(1), \ldots, S(t-2)$ are all the previous positions. Hence, $P[S(t) = h|S(0), \ldots, S(t-1)] = P[S(t) = h|S(t-1)]$.
- A simple Random walk $S(t)$ is spatially homogeneous $P[S(t) = h|S(0) = a] = P[S(t) = h + b|S(0) = a + b]$.
- A simple Random walk $S(t)$ is temporally homogeneous $P[S(t) = h|S(0)] = P[S(t+m) = h|S(m)]$ implying it's a stationary process.
- $S(t) = \sum_{j=0}^{t-1} X(j) + X(t) = X(0) + \cdots + X(t)$, where$X(0), \ldots, X(t)$ are iids.
- A simple random walk is a stationary process $F_{X_0,\ldots,X_t}(x_0, \ldots, x_t) = F_{X_\tau,\ldots,X_{t+\tau}}(x_0, \ldots, x_t)$.

Time Series as a Random Walk

- The size of each step in the random walk $S(t)$ determines the type of random walk.
- For example, if $X(t) \epsilon [-1, 1]$ for $t \epsilon \mathbb{N}$ with $P[X(t) = 1] = p$ and $P[X(t) = -1] = 1 - p$, then S is a discrete-value random walk, $S(t) \epsilon \mathbb{Z}$ called a Bernoulli Random Walk.
- Similarly, a random walk with $S(t) \epsilon \mathbb{R}$ and $t \epsilon \mathbb{R}$ is called a continuous-time random walk with jumps on a real line. An example is an AR-1 stationary process with Gaussian Distribution at each step, $S_{t+1} = S_t + \epsilon_t$ where $\epsilon_t \sim \mathcal{N}[\mu, \sigma^2]$, and $S_0 = \epsilon_0$.
- Furthermore, a random walk with $S(t) \epsilon \mathbb{R}^n$ and $t \epsilon \mathbb{R}$ is called a continuous-time random walk in n-dimensional space.
- If each step $X(t) \ \sim \mathcal{N}[0, \tau]$, $t \epsilon \mathbb{R}$, and $\tau = \Delta t$, it's called a Wiener process, and the continuous-value random walk is called the Brownian motion, where $S(t) \epsilon \mathbb{R}$.