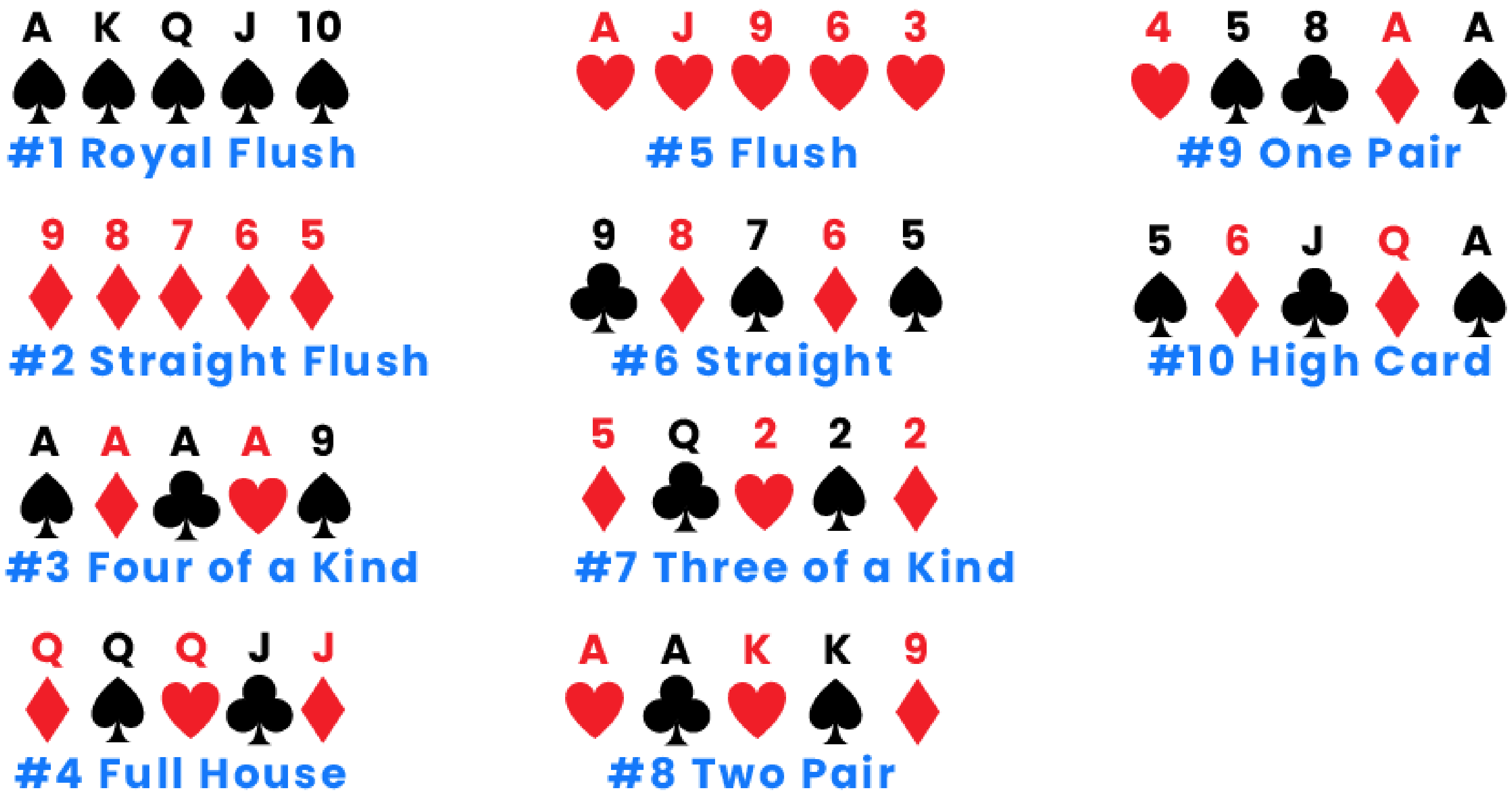
Lecture 1:

Probability and Indicator

MSDM 5058

Prepared by S.P. Li

The Game of Poker



Poker Hand Rankings

List of patterns in poker hands. There is roughly an even chance that any five-card hand will form one of the patterns on the list below. The chances of getting a specific royal flush are 1 in 2,598,960 (= 52!/((52-5)!5!)) hands. You are therefore five times more likely to get struck by lightning than get the same hand twice!

Pattern	Number of possibilities
Royal flush	4
Straight flush	36
Four of a kind	624
Full house	3,744
Flush	5,108
Straight	10,200
Three of a kind	51,168
Two pairs	123,552
Two of a kind	1,098,240

Quick Review of Some Basics in Probability Theory

Objective of Probability Theory

- Attach real numbers $\mathbf{0} < \boldsymbol{p} < \mathbf{1}$
- $\boldsymbol{p}$ = probabilities to the results of chance experiments
- Two events are ***mutually exclusive*** if the occurrence of one precludes the occurrence of the other.
- A set of events is said to be ***exhaustive*** if one of them must occur in the experiment.
- The ***null*** event cannot happen, and the **certain** event must happen in the experiment.

Quick Review of Some Basics in Probability Theory (cont'd)

Random Variable

A random variable is a rule that assigns a real number to every outcome of a chance experiment.

For example, in flipping a coin,

assigning $X = +1$ to the occurrence of a head

and $X = -1$ to the occurrence of a tail

constitutes the assignment of a discrete-valued random variable

Quick Review of Some Basics in Probability Theory (cont'd)

Transformation of Random Variable

If $Y = g(x)$ is a monotonic function, then the transformation of the random variable from x to y gives a corresponding transformation of the *pdf*

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|_{x=g^{-1}(y)}$$

Here $f_Y(y)$ is the new *pdf* for the transformed variable y

Quick Review of Some Basics in Probability Theory (cont'd)

Cumulative Distribution Function

Cumulative distribution function (*cdf*)of random variable X is defined as the probability that $X < x$ where x is the real number X assuming

$$F_X(-\infty) = 0\,; \qquad F_X(\infty) = 1\,; \qquad 0 \leq\ F_X(x) \leq 1$$

$F_X(x)$ is a non-decreasing function of its argument.

Discrete random variables have discontinuous *cdf*'s

Continuous random variables have continuous *cdf*'s.

Quick Review of Some Basics in Probability Theory (cont'd)

Probability Density Function (pdf)

The probability density function is $f_X(x)$ defined as

$$f_X(x) = \frac{dF_X(x)}{dx}$$

$$F_X(x) = \int_{-\infty}^{x} dx' f_X(x')$$

Note that $0 \leq F_X(x) \leq 1$ and $\int_{-\infty}^{\infty} f_X(x)dx = 1$

$f_X(x)dx$ = probability of the random variable X lying between $(x, x + dx)$

Quick Review of Some Basics in Probability Theory (cont'd)

Probability Density Function (pdf)

Example 1

A fair die with M faces has $X = \{1, 2, \ldots, M\}$ and $p(i) = 1/M$ for all $i \in \{1, \ldots, M\}$. The average of x is $\boldsymbol{E}(X) = (1 + \ldots + M)/M = (M + 1)/2$.

Example 2

Gaussian variable. A continuous variable $X \in R$ has a Gaussian distribution of mean m and variance σ^2 if its probability density is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

where $\boldsymbol{E}(X) = m$ and $\boldsymbol{E}(X - m)^2 = \sigma^2$.

Quick Review of Some Basics in Probability Theory (cont'd)

Joint cdf, pdf   *Marginal cdf, pdf*

- The joint *cdf* of two random variable X and Y is the probability $F_{XY}(x, y)$ that $X \leq x$ and $Y \leq y$
- The joint *pdf* of two random variable X and Y is

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

- Marginal *cdf* of X alone = $F_{XY}(x, y = \infty)$
- Marginal *pdf* of X alone = $m_X(x) = \int_{y=-\infty}^{y=+\infty} f_{XY}(x,y)dy$

Statistical Independent Events

Two statistically independent random variables have joint *cdf*'s and *pdf*'s that factors into the respective marginal *cdf*'s or *pdf*'s.

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

$$F_{XY}(x,y) = F_X(x)F_Y(y)$$

Quick Review of Some Basics in Probability Theory (cont'd)

Relation between Events

Given two events A and B,

"**A or B or both**" is denoted as $\mathbf{A \cup B}$

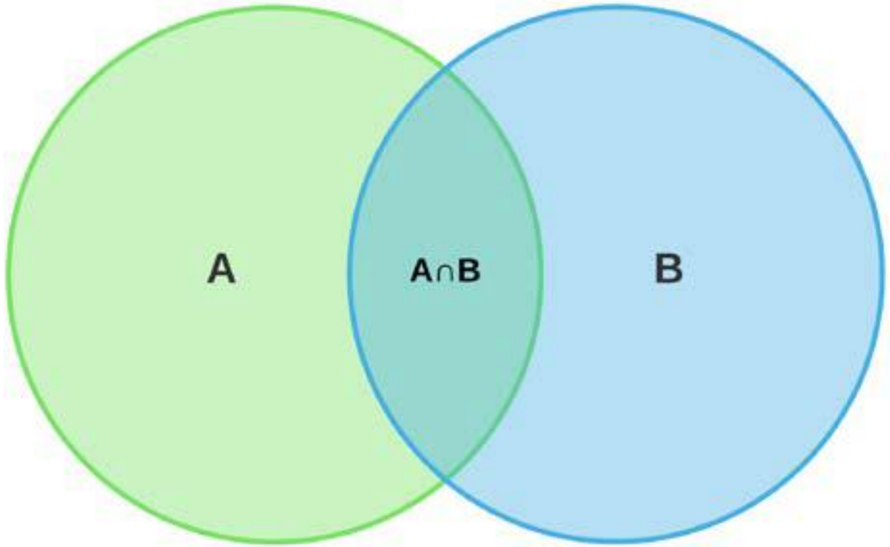
"**both A and B**" is denoted as (A and B) or $\mathbf{A \cap B}$

"**Not A**" is denoted as $\bar{A}$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

= ***total colored area***

= ***Green + Blue – Intersection.***



Law of Total Probability

If $\{B_i : i = 1,2,3,\ldots\}$ is a finite or countably infinite partition of a sample space and each event B_i is measurable, then for any event of the same probability space

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Sequential Conditional Probability

$$P(A_1 \cap A_2) = P(A_2|A_1)P(A_1)$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_3|A_2A_1)P(A_2|A_1)P(A_1)$$

..... = ...

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_n|A_{n-1} \ldots A_3A_2A_1)P(A_3|A_2A_1)P(A_2|A_1)P(A_1)$$

Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example: Two fair dice are rolled. What is the conditional probability that at least one lands on 6 given that the dice land on different numbers?

Answer: A = {at least one of them is 6}; B = {two dice show different numbers}

$$P(A \cap B) = \frac{10}{36}; P(B) = 1 - P(B^c) = 1 - \frac{6}{36}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{10}{36}}{\frac{30}{36}} = \frac{1}{3}$$

Bayes' Formula

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$$

$$\longrightarrow \quad P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Example: An ectopic pregnancy is twice as likely to develop when the pregnant woman is a smoker as it is when she is a nonsmoker. If 32% of women of childbearing age are smokers, what percentage of women having ectopic pregnancies are smokers?

Bayes' Formula (cont'd)

Answer:

Let E be the event of ectopic pregnancy, S be the woman smokers, and S^c be the non-smokers. Then,

$P(S) = 0.32$ and $P(S^c) = 0.68$. Let $P(E|S^c) = a$, then $P(E|S) = 2a$.

$$P(E) = P(E|S)P(S) + P(E|S^c)\,P(S^c) = 2a * 0.32 + a * 0.68 = 1.32a$$

Using Bayes' formula, $P(S|E)P(E) = P(E|S)P(S)$, we have

$$P(S|E) = \frac{P(E|S)P(S)}{P(E)} = \frac{0.64a}{1.32a} = \frac{16}{33} = 0.4848$$

Bayes' Formula: A Diachronic Interpretation

A way to update the probability of a hypothesis, *H*, in light of some body of data, *D*.

$$P(H|D)P(D) = P(H \cap D) = P(D|H)P(H)$$

$$\longrightarrow \qquad P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

- $P(H)$ is the probability of the hypothesis before we see the data, called the ***prior***.
- $P(H|D)$ the probability of the hypothesis after we see the data, called the ***posterior***.
- $P(D|H)$ is the probability of the data under the hypothesis, called the ***likelihood***.
- $P(D)$ is the probability of the data under any hypothesis, called the ***normalizing constant***.

Summation and Averaging

The n-th moment of X:

$$E(X^n) = \langle X^n \rangle = \int_{-\infty}^{\infty} f_X(x) x^n$$

The n-th central moment of X:

$$\langle (X - E(X))^n \rangle$$

The first moment is the mean = $E(X)$

The second central moment is the variance = Var (X) = $\sigma_X{}^2$

Summations and averaging can be interchanged

The variance of the sum of two statistically independent random variables is the sum of the respective variance.

Characteristic Function

The characteristic function $M_X(iu)$ of a random variable X that has pdf $f_X(x)$ is the expectation of e^{iux}

$$M_X(u) = \int_{-\infty}^{\infty} f_X(x) e^{iux}\, dx$$

$$f_X(x) = \int_{-\infty}^{\infty} M_X(x) e^{-iux}\, du$$

pdf and its characteristic function form a Fourier Transform pair

Note that the characteristic function M is in general a complex function.

The n-th moment of X can be obtained from its Characteristic function by differentiation

$$(-i)^n \times \left(\left. \frac{d^n M_X(u)}{d^n u} \right|_{u=0} \right) = \int_{-\infty}^{\infty} f_X(x) x^n \; = \langle X^n \rangle$$

Characteristic Function for sum of Independent Random Variables

If $Z = X + Y$, where X and Y are independent, then

$$M_Z(u) = M_X(u)\, M_Y(u)$$

Mathematical result (exercise for those who know Fourier Transform), we can show that the *pdf* of Z satisfies this relation

$$f_Z(v) = \int_{-\infty}^{\infty} f_X(x-u) f_Y(u)\, du$$

This is called the convolution of the *pdf*'s of X and Y.

Covariance

The covariance of two random variables X and Y is

$$\Gamma_{XY}(x,y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, f_{XY}(x,y)(X - \langle X \rangle)(Y - \langle Y \rangle)$$

The correlation coefficient for X and Y is defined as

$$\rho_{XY} = \frac{\Gamma_{XY}}{\sigma_X \sigma_Y} \text{ with } \sigma_X = \sqrt{Var(X)} \text{ and } \sigma_Y = \sqrt{Var(Y)}$$

Range of Correlation Coefficient

One can show mathematically that $-1 \leq \rho_{XY} \leq 1$

We say that the variables X and Y are positively (negatively) correlated if their correlation coefficient is positive (negative).

If correlation coefficient is close to zero, we say that the variables X and Y are not correlated

Statistical Independence and Correlation

If X and Y are statistically independent, one can show that the correlation coefficient between X and Y is zero

However, if correlation coefficient between X and Y is zero, X and Y need not be statistically independent!!!

Example 1: Nonlinearity

Zero correlation will indicate no linear dependency, however it won't capture *non-linearity*.

Consider a random variable X uniformly distributed over [-1,1] and X has zero mean. Introduce the associated random variable $Y=x*x$

Correlation between X and Y is zero but they are not independent

Example 2: Car Traffic

Consider a car travelling at velocity V between city A and B. V is a random variable between $[-50,+50]$ miles an hour with equal probability, average velocity 0.

Moving towards B, velocity $V > 0$; Moving towards A, velocity $V < 0$

The kinetic energy $K = \frac{1}{2} mV^2$ where m is the mass of the vehicle. Note that $V = \pm 50$, $K = 1250\ m$

What is the correlation between V and K?

Since the mean velocity is zero and all velocities are equally likely, (we see that KV when $V = -50$ is $-62500\ m$, and when $V = +50$, $KV = 62500\ m$ and because negative velocities occur just as much as positive velocities)

Thus, the correlation between V and K is zero On the other hand, V and K are not independent, we see that the joint pdf of (V,K) cannot be factorized as the pdf of V and pdf of K

Time Series Data Analysis

For application in financial time series, we need to consider if there is any way we can make prediction as we like to know if one can find pattern in the past that works for future data. This requires us to employ the idea of independence and correlation.

A time series of a random variable such as stock price $X(t)$ has ***serial dependence*** if the value at some time t in the series is statistically dependent on the value at another time t'.

A time series is ***serially independent*** if there is no dependence between any pair. Intuitively, if a time series is ***serially independent***, we cannot make any prediction.

Autocorrelation Function

For financial time series, the data are time ordered.

We can test if this time series has extra information in its temporal patterns (i.e. if the data is serially dependent).

The autocorrelation function $R(\tau)$ is useful to find these patterns as it tells you the correlation between points separated by various time lags.

Consider a time indexed set of numbers, such as the stock price $x(t)$ as a function of time, we define the time autocorrelation function with delay τ as

$$R(\tau) = \langle x(t)x(t+\tau)\rangle = \lim_{T\to\infty} \int_{-\infty}^{+\infty} x(t')x(t'+\tau)dt'$$

Stationary process

In mathematics, a stationary process is a process whose unconditional joint probability distribution does not change when shifted in time. Consequently, parameters such as mean and variance also do not change over time. In real life, stationary process is rare and usually is a simplifying assumption for mathematical analysis.

Stationary Time Series

If a time series $\{X(t)\}$ is stationary, then statistical dependence between the pair

$$(X(t), X(s = t + \tau))$$

would imply that there is statistical dependence between all pairs of values at the same lag $\tau = s - t$.

Analysis of a Finite Discrete set of Time Series data

In real data, $\{X(t)|t_i > t_j, \text{if } i > j, i \in (0, 1, 2, \dots, N)\}$,

N denotes the number of data which is finite and the value of stock price $X(t)$ is discrete, indexed by the time t *in increasing time starting at the time origin* t_0.

How should one compute the autocorrelation and test if one can perform datamining of the temporal patterns hidden in the data?

Mean Reverting Time Series Data

Since stationarity is an assumption underlying many statistical procedures used in time series analysis, non-stationary data is often transformed to become stationary with some tricks, ***sometimes*** successfully.

A common cause of violation of stationarity is a trend in the mean.

In finance, we often assume ***mean reversion of the stock price***, which is the assumption that a stock's price will tend to move to the average price over time.

This implies that there is some kind, often unknown, but in principle determinable trend of the mean.

Trending a Mean Reverting Time Series

Many real time series is not a trend stationary process, e.g., the so-called cyclo-stationary process, which is a stochastic process that varies cyclically with time.

However, a common assumption for elementary financial time series analysis to assume that the time series is mean reverting, which means it is a trend stationary process.

In this case, sudden shocks have only transitory effects after which the variable tends toward a deterministically evolving (non-constant) mean.

A trend stationary process is not strictly stationary, but can easily be transformed into a stationary process by removing the underlying trend, i.e. subtracting its mean.

Daily Rate of Return of a Discrete Data Set

There are many common ways to compute the mean of a set numbers in finance. We first introduce the daily return and log return.

Consider your chosen financial time series data $\{s(0), s(1), \ldots, s(N)\}$ and form the new series of $\{x(1), x(2), \ldots, x(N-1)\}$.

$$x(t) = \frac{s(t) - s(t-1)}{s(t-1)}$$

Note that N is the number of points in your time series, which should be large. This series of x is called the daily rate of return in finance.
People also use the **log return** as the daily rate of return, $x(t) = \log(s(t)/s(t\text{-}1))$.
In any case, the daily rate of return is a number, or fraction of change of the stock price $s(t)$.

General Moving Average for the Mean

Given a finite set of $N + 1$ discrete data $\{s(t)|t_i = t_0 + i\varepsilon, i \in (0, 1, 2, \dots, N)\}\}$
Here $\varepsilon = \Delta t$ is the time difference between data and is constant.
We define the Moving Average $M(a = m\varepsilon > 0;\ t_j = t_0, j \geq m > 0;\ \tau)$,
for $(s(t))$ over the subset value of m data points before
It is important to define the number m of past values to be included in the computation of moving average with weighting factor W and time delay

$$M\left(a = m\varepsilon > 0;\ t_j = t_0, j \geq m > 0;\ \tau\right) = \sum_{i=1}^{m} s(t_i)W(t_i) \ \text{ with } m > 0$$

and $W(t_i) = \dfrac{e^{-\left(t_j - t_i\right)/\tau}}{Z}$, $Z \equiv \sum_{i=1}^{m} e^{-\left(t_j - t_i\right)/\tau}$, $t_i = t_j - (i-1)\varepsilon$

Simple Moving Average for the Mean

If we set τ = ∞ for the data $\{s(t)|t_i = t_0 + i\varepsilon, i \in (0, 1, 2, \dots, N)\}$
Then,

$$M(a = m\varepsilon > 0; t_i = t_0 + m\varepsilon, \tau = \infty) = \sum_{i=1}^{m} s(t_i)W(t_i) = \frac{1}{m}\sum_{i=1}^{m} s(t_i)$$

with m > 0, $W(t_i) = \dfrac{e^{-(t_j - t_i)/\tau}}{Z}$, $Z \equiv \sum_{i=1}^{m} e^{-(t_j-t_i)/\tau}$, $t_i = t_j - (i-1)\varepsilon$

The meaning of τ = ∞ is that we assume equal weight of the past values of $s(t)$ in computing the average as the weighting factor $W(i)$ is the same $1/m$. It also implies that the data set has long memory.

*In real data, usually the memory is short with τ << ∞

EMA and SMA for the Mean

An exponential moving average (EMA) is a moving average (MA) that places a greater weight and significance on the most recent data points. Mathematically, it takes the form

$$EMA = \text{Closing price} \times \lambda + EMA\ \text{(previous day)} \times (1 - \lambda)$$

where λ *EMA* reacts more significantly to recent price changes than a simple moving average (*SMA*), which applies an equal weight to all observations in the period.

- *EMA* is a moving average that places a greater weight and significance on the most recent data points.
- *EMA* is used to produce buy and sell signals based on crossovers and divergences from the historical average.

Moving Average with different periods

If we set $\tau = \infty$ for the data $\{s(t)|t_i = t_0 + i\varepsilon, i \in (0, 1, 2, \dots, N)\}$

We can consider different periods $a_k = m_k\varepsilon$ to get different moving averages for the mean

$$M_k(a_k = m_k\varepsilon > 0; t_k = t_0 + m_k\varepsilon; \tau)$$

It is common in the technical analysis of financial data to consider different $a_k = m_k\varepsilon$ such as m_k = 10, 20, 50, 100, 150, 250 which corresponds to short-term (10 for 2 weeks, 20 for 1 month) and long-term (250 for one year) behavior of the moving averages of the mean.

There are also simple rules such as MACD used in these analyses.
(Moving Average Convergence / Divergence指數平滑異同移動平均線)

Example: HSBC stock (0005)

Stock price of the Hong Kong Shanghai Bank (stock code 0005)

MACD

Moving Average Convergence Divergence (*MACD*) is a trend-following momentum indicator that shows the relationship between two moving averages of a security's price.

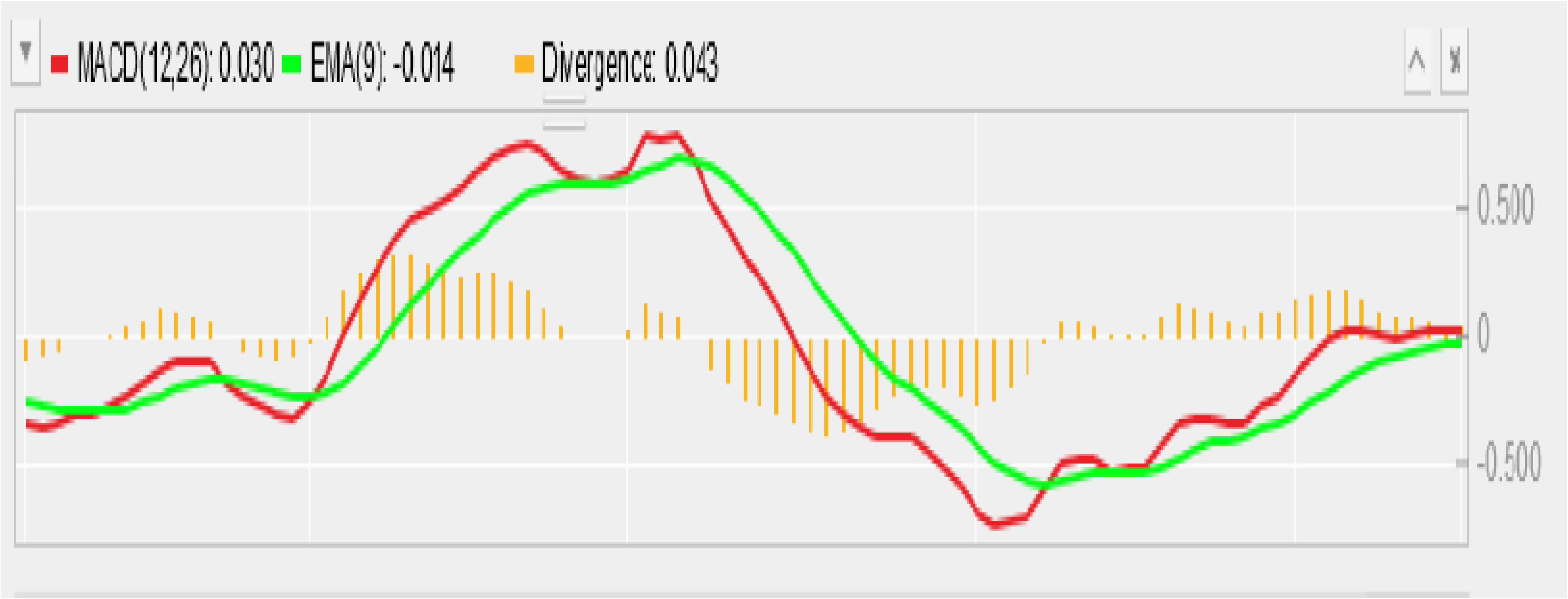
MACD is calculated by subtracting the *26-period Exponential Moving Average* (*EMA*) from the *12-period EMA*, which yields what we called the ***MACD line***.

A ***nine-day EMA of the MACD*** called the "**signal line**", is then plotted on top of the MACD line, which can function as a trigger for buy and sell signals.

Traders may buy the security when the MACD crosses above its signal line and sell or short the security when the MACD crosses below the signal line

Example: HSBC stock (0005)

MACD of HSBC



Convergence

Convergence describes the phenomenon of the futures price and the cash price of the underlying commodity or stock moving closer together over time.

Convergence happens because, theoretically, an efficient market will not allow something to trade for two prices at the same time.

We will discuss the efficient market hypothesis later in the course.

What is interesting is divergence!

We now discuss divergence of *MACD*

Signal line and MACD line

MACD is positive whenever the 12-period EMA (blue) is above the 26-period EMA (red) and is negative when the 12-period EMA is below the 26-period EMA.

The base line is defined by the difference of the blue line from the red in the top figure

In the bottom figure, we see the subtraction to yield the blue MACD line and the signal line is the 9 days EMA orange line

Divergence and crossover in MACD

MACD line triggers technical signals when it crosses above (to buy) or below (to sell) its ***signal line***. The speed of crossovers is also taken as a signal of a market is overbought or oversold.

- When *MACD* falls below the signal line, it is a bearish signal which indicates that it may be time to sell,
- When *MACD* rises above the signal line, the indicator gives a bullish signal, which suggests that the price of the asset is likely to experience upward momentum.
- When *MACD* rises or falls rapidly, it is a signal that the stock is overbought or oversold and will soon return to normal levels.

Positive or negative crossovers, divergences, and rapid rises or falls can be identified on **the histogram of the difference of MACD line and the signal line.**

Other Indicators:

- Bollinger Bands
- RSI (Relative Strength Index)
- ………