Clip 1: An Introduction to Pricing

Mark-Jan Boes

September, 2020

This course is about finding the **no-arbitrage price** of **derivatives** contracts.

No-arbitrage pricing means that prices are determined in such a way that there is **no free lunch** in the market.

Simple example:

- I toss a coin: if heads comes up I give you EUR 10 and if tails comes up I give you EUR 20
- I offer you this under the condition that you pay me EUR 5 before tossing the coin

You will enter into this deal, because whatever the outcome of the toss coin is you will receive more than you paid.

Hence, a price of EUR 5 for this particular deal creates arbitrage opportunities.

Björk defines an **arbitrage portfolio** as a portfolio h with the properties:

$$
\begin{aligned}
&V_0^h = 0 \\
&V_1^h > 0 \text{ with probability } 1.
\end{aligned}
$$

An arbitrage portfolio is thus a deterministic money making machine.

We can combine the example with the definition as follows:

- You borrow EUR 5 and give this to me as price for the bet
- Your portfolio value is therefore equal to EUR 0
- If heads comes up: you receive EUR 10 from me and you give back EUR 5 to your lender
- If tails comes up: you receive EUR 20 from me and you give back EUR 5 to your lender
- So, for all possible outcomes of the coin toss you end up with some money while you started with nothing

In the definition, the no-arbitrage price should be at least EUR 10.

Is there something more that we can say about the no-arbitrage price of this bet?

Perhaps you remember the following two concepts of earlier courses:

- The Law of One Price
- Equilibrium pricing

In both approaches we look at a broader perspective: a market in which multiple financial instruments are traded under the assumption that the market is frictionless and does not allow for arbitrage opportunities.

The Law of One Price says that if portfolios or instruments have the same payoff these portfolios or instruments should have the same price.

In our example:

- Suppose we have a market where the following bet is traded: if heads comes up the seller pays EUR 5 to the buyer and if tails comes up the seller pays EUR 10 to the buyer
- In the market this bet trades at a price of EUR 7
- What is the no-arbitrage price of the bet we discussed earlier using the principle of the Law of One Price?

Obviously, the answer to this question is EUR 14.

The bet that is traded in the market provides exactly half of the payoff of the bet that I offered you at the start.

Hence, buying two bets in the market gives exactly the payoff of the bet I offered you: then the no-arbitrage price of my offer should be twice the market price.

Suppose the price of my offered bet would be EUR 12, what could you do?

- Act as a seller in the market: sell 2 bets for a combined price of EUR 14
- Use the money received to buy one bet from me against EUR 12
- A market referee tosses the coin
- No matter what the outcome is: you'll end up with EUR 2

This principle is applied a lot when pricing derivatives: the no-arbitrage of derivatives can in quite a few cases be derived from the prices of base instruments like stocks, bonds and cash.

The pricing method is also known as the **replication method**: construct a portfolio of base instruments that **replicates** the payoff of the derivative perfectly (in each possible realisation of the world).

The Law of One Price tells us that the price of strategy and the derivative instrument should be exactly the same, otherwise arbitrage opportunities are present.

Hence, with the **replication method** we can (often) calculate the no-arbitrage price of derivatives **conditional on** the knowledge of market prices of other instruments.

But: where do those market prices come from? And how can we be sure that these prices are fair?

That has already been illustrated to you in earlier courses: one example of a pricing model that we treat extensively in Finance courses is the **Capital Asset Pricing Model**.

Most of you will think about the following when I mention **Capital Asset Pricing Model**:

$$\mathbb{E}(R^i) = R^f + \beta^i \mathbb{E}(R^M - R^f).$$

The expected return R on a stock i is fully determined by its β, β measures the exposure of the stock toward the market portfolio.

The idea of the CAPM is that only exposure to the risk in the market portfolio M is compensated where compensation means a higher expected return.

In factor language: there is one factor (the market portfolio) that drives the cross-section of expected stock returns.

But: do you still remember where this formula comes from?

To refresh your memory:

- Firms need capital for investment
- One way of raising capital is by issuing shares, i.e. shares are **supplied** to the market
- On the other hand we have investors who want to bring capital to the market: they have a **demand** for shares
- Prices will settle there where demand for stocks **exactly** meets supply for stocks: these are the **equilibrium** prices
- If prices are low relative to expected payoffs, the demand of investors will be higher than supply with the consequence that prices will rise

From a modelling perspective:

- The CAPM is a 1-period model
- **All** investors in the market have the **same view** on the probability distribution of all risky payoffs (this is quite an important and heroic assumption!)
- Investors care only about the mean and variance of their investment portfolio, i.e. they perform a mean-variance analysis to find their demand for the risky assets
- Exact demand depends on risk aversion of the investors: the more risk averse the less eager the investor is to the invest in risky assets
- Prices converge to a level where the demand that follows from the mean-variance model exactly meets market supply
- The composition of the risky part of each investor's portfolio is exactly the same: this is the **market portfolio**.

For more details and extensions I refer to earlier courses and to the course Asset Pricing.

From a no-arbitrage perspective we could say that the price of a stock should be at least 0 because the worst that can happen that investors lose the full investment (payoff equal to zero).

Exact stock prices (which will be larger than 0) follow from an **equilibrium model** e.g. the **Capital Asset Pricing Model**.

Two comments:

- Pricing critically depends on model assumtions which are pretty strong in the case of the **CAPM**
- The **Law of One Price** is not useful in this case because all stocks in the market are different

A few points that summarize this knowledge clip:

- Purpose of the course is to derive the **no-arbitrage price** of derivative instruments
- In this clip we focused on what **arbitrage pricing** means
- We've tried to place the concept of **arbitrage pricing** in the context of pricing methods you've seen before: the Law of One Price and Equilibrium pricing
- We established that the Law of One Price can only be applied if you have access to knowledge on (no-arbitrage) prices of other assets in the market
- This is a condition that is typically met when we price derivatives: the replication method is an often applied method (also in this course)
- Equilibrium pricing is usually applied when pricing stocks

Clip 2: Derivatives contracts

Mark-Jan Boes

September, 2020

This course is about finding the **no-arbitrage price** of **derivatives** contracts.

No-arbitrage pricing means that prices are determined in such a way that there is **no free lunch** in the market.

Björk defines an **arbitrage portfolio** as a portfolio h with the properties:

$$V_0^h = 0$$
$$V_1^h > 0 \text{ with probability } 1.$$

An arbitrage portfolio is thus a deterministic money making machine.

We covered no-arbitrage pricing in knowledge clip 1.

Purpose of this second knowledge clip is to refresh your memory on financial derivatives contracts.

Derivatives:
The value of a financial derivative contract depends on (i.e. is **derived** from) a particular economic variable or other financial asset.

Today we will briefly look at:

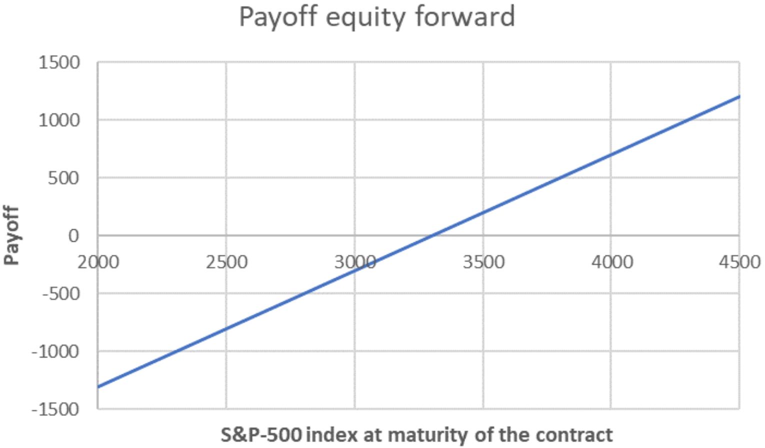
- Equity forwards
- Currency forwards
- European call options on equity (indices)
- European put options on equity (indices)

Equity forwards:

- The long position ('the buyer') has **the obligation** to buy the underlying stock (index) at a future moment for a price (a.k.a. the reference price) that the buyer and the seller agree upon today
- The reference price is usually set in such a way that there is no exchange of cash at the start, i.e. the contract value is 0 at inception

Example: if you have a positive view on the development of the S&P-500 index (a broad US stock index) you can take a long position in a 3-months forward contract with the the S&P-500 index as underlying and reference price (for instance) USD 3,300.

What is the payoff of this contract at the maturity date?



The idea is simple:

- If at maturity date the value of the S&P-500 index is equal to 4,300
- Then the holder of the equity forward is obliged to buy the S&P-500 index of the seller for 3,300
- The buyer sells the S&P-500 index in the market for 4,300
- The buyer is left with USD 1,000 in cash which is then the payoff of the contract

In mathematical terms, the payoff V of a long position in the equity forward is given by

$$V(T) = S(T) - K,$$

where S denotes the underlying value, T the maturity date of the contract, and K the reference price.

The main question is: **what is the no-arbitrage price of an equity forward**?

We will treat this question in the third knowledge clip.

Suppose a Dutch company (EUR as base currency) expects to receive USD 1,000,000 in three months from now.

This company is exposed to FX-risk.

To illustrate:

- Suppose that the EUR / USD FX-rate is now 1.250 (meaning: 1 Euro trades for 1.250 US dollars)
- Hence, if the FX-rate won't change during the next three months the company can convert USD 1,000,000 into EUR 800,000 then
- However, if the dollar depreciates versus the Euro, e.g. the FX-rate becomes 1.333, then the EUR-income will only be EUR 750,000

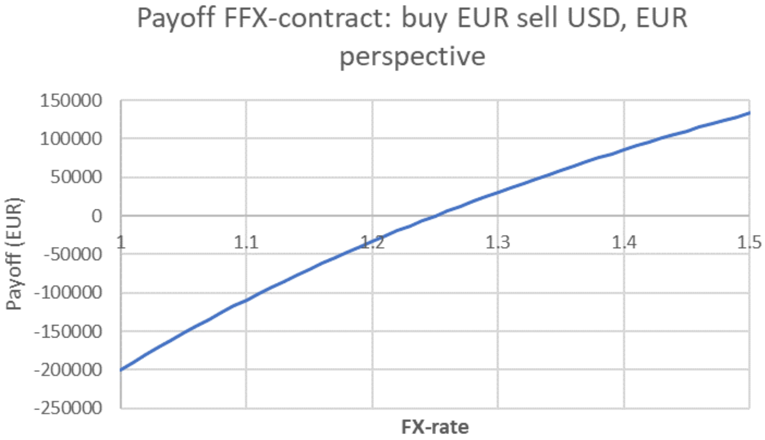
It would therefore be very helpful for the company if it could enter into a contract in which the company could sell USD 1,000,000 in three months from now against a predetermined number of Euros (determined at the trading time of the contract).

Entering in such a contract would take away **all currency risk**.

The FX-forward contract (FFX) is exactly providing this:

- Two parties agree to exchange money in different currencies at a future point in time.
- It is an **obligation** for both parties
- The currencies and amounts are set at the start of the contract, typically in such a way that there is no exchange of cash at the start of the contract

Suppose that the contract would trade at the current spot rate of 1.250, how would the payoff look like?



In order to see this:

- suppose that the EURUSD FX-rate is 1.500 at maturity of the contract
- in order to fulfill obligations of the contract, the 'EUR-buyer in the FFX-contract' borrows EUR 666,666.67 and converts this to USD 1,000,000
- the USD 1,000,000 are delivered to the counterparty in the FFX-contract
- the receipt is EUR 800,000
- the 'EUR-buyer in the FFX-contract pays off his debt of EUR 666,666.67 and is left with EUR 133,333.33

In mathematical terms, the payoff V of a long position in the FX-forward contract is given by (buy EUR, sell USD from a EUR perspective):

$$V(T) = N_T^{USD}\left(\frac{1}{K^{FFX}} - \frac{1}{FX_T}\right),$$

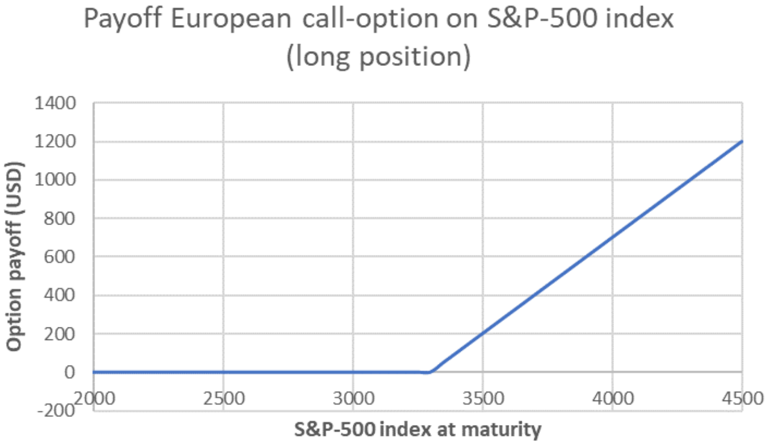
where

- T is the maturity date of the contract
- N the amount in dollars (to be paid in this example)
- K^{FFX} is the reference FX-rate of the contract (determined at the start of the contract)

The third contract that we'll discuss is the **European call option on stocks**:

- The buyer of the option (the long position) has **the right**, at the maturity of the contract, to buy the underlying stock at **the exercise price** K
- The exercise price is part of the contract conditions that are determined at trade moment of the contract
- 'Option' refers to the characteristic that the holder has the right, i.e. **not** an obligation, to buy the underlying
- 'European' refers to the characteristic that the holder **can only exercise** at the maturity date: in American options the holder can also exercise before the maturity date

Payoff profile of a 3-months European call option on the S&P-500 index with exercise price 3,300.



The logic is simple:

- If the S&P-500 index is 4,500 at maturity
- Then the buyer exercises the right to buy the S&P-500 index against 3,300
- He sells the acquired index in the market against 4,500 which leaves 1,200 in cash for him

It is obvious that from a no-arbitrage perspective, the call option should trade at a positive value otherwise the profit and loss (**P&L**) for the investor would always be positive at maturity.

From a mathematical perspective the call option payoff, V is given by:

$$V(T) = \max(S(T) - K, 0),$$

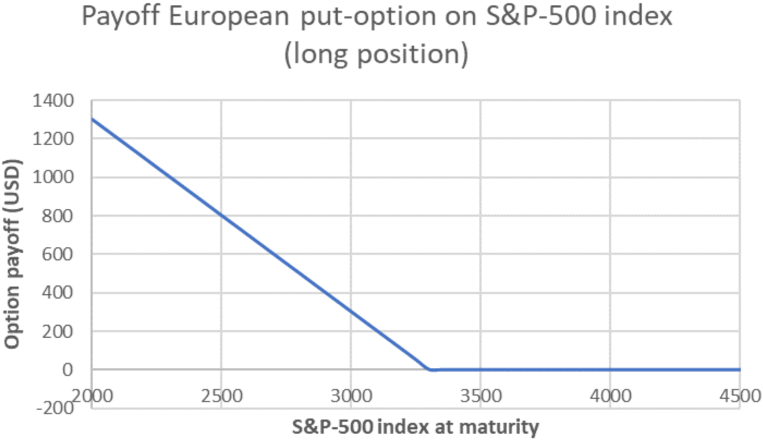
where

- T is the maturity date of the option
- S is the underlying value (in this case stocks) of the option
- K is the strike price of the option

The final contract that we'll discuss is the **European put option on stocks**:

- The buyer of the option (the long position) has **the right**, at the maturity of the contract, to sell the underlying stock at **the exercise price** K
- The exercise price is part of the contract conditions that are determined at trade moment of the contract
- 'Option' refers to the characteristic that the holder has the right, i.e. **not** an obligation, to buy the underlying
- 'European' refers to the characteristic that the holder **can only exercise** at the maturity date: in American options the holder can also exercise before the maturity date

Payoff profile of a 3-months European put option on the S&P-500 index with exercise price 3,300.



The logic is similar to the European call option:

- If the S&P-500 index is 4,500 at maturity
- Then the buyer of the put option won't do anything
- It is not sensible to exercise the option because the underlying stock can be sold in the market for a much higher price than the exercise price of the option
- Therefore the payoff of the option is zero in this case

From a mathematical perspective the put option payoff, V is given by:

$$V(T) = \max(K - S(T), 0),$$

where

- T is the maturity date of the option
- S is the underlying value (in this case stocks) of the option
- K is the strike price of the option

In this knowlegde clip we have discussed the **payoff profiles** of four financial derivatives contracts:

- Equity forwards
- Currency forwards
- European call options on equity (indices)
- European put options on equity (indices)

Next step is to determine the **“correct” price** for these derivatives.

Clip 3: An Essential Observation with respect to Pricing

Mark-Jan Boes

September, 2020

Part I

Pricing: Essentials

This course is about finding the **no-arbitrage price** of **derivatives** contracts.

No-arbitrage pricing means that prices are determined in such a way that there is **no free lunch** in the market.

Björk defines an **arbitrage portfolio** as a portfolio h with the properties:

$$V_0^h = 0$$
$$V_1^h > 0 \text{ with probability } 1.$$

An arbitrage portfolio is thus a deterministic money making machine.

In clip 1 we have seen that the **The Law of One Price** is a powerful tool to calculate the no-arbitrage price of a financial derivative: if you are able to setup a strategy of **market** instruments that replicates the payoff profile of the financial derivative **exactly** then you know that the price of the replicating portfolio and the financial derivative should be the same.

But what if the underlying risk in a financial instrument is **not traded** in a market, what can you then?

You might know from Corporate Finance courses that the decision to invest of a project is based on making a **projection** of future cash-flows and **discounting** these to today.

Is this something that we can apply? Let us look at an example.

Suppose you have a job. Your employer pays you EUR 1,000 per month.

Payday has arrived and you have EUR 1,000 available. Suppose you have two investment opportunities:

- You keep the money in your bank account: this delivers you 0% return for the next year (which is pretty realistic nowadays...)
- Mark-Jan drops by and says: 'give me the money today and I will deliver you EUR 1,111.11 in one year from now'

The only problem with the second option is that the probability that Mark-Jan goes bankrupt is 10%.

Simple question:

What would you do with your EUR 1,000?

In this example we see the two of the most important (and related) questions from the finance profession at work:

- How do I sensibly allocate my assets across the available investment opportunities?
- How do I price financial assets in a sensible way?

Both investment opportunities can be regarded as an investment in a bond or loan:

- Posting EUR 1,000 in your bank account is a risk free investment (guaranteed by the deposit guarantee system)
- Giving EUR 1,000 to Mark-Jan is similar to investing in a corporate bond

Provided that the 1-year risk free interest rate is 0%, pricing of the risk free investment is simple: the asset **should** earn the risk free rate.

Hence,

- The expected payoff is EUR 1,000 in one year
- Discount this expected payoff with the correct discount rate (0%)
- End result: value equals EUR 1,000

Note that if the bank would give you a little bit more than the risk free rate, your investment will be worth more than EUR 1,000 today.

But what about the price of the second investment opportunity?

Let us first look at the expected payoff V at time $t = 1$:

$$\mathbb{E}_0(V_1) = 90\% \times 1,111.11 + 10\% \times 0 = 1,000.$$

Hence, the **expected payoff** of this investment is exactly the same as of the first investment opportunity.

OK, does that mean that both investments should have the same price as well?

That truly depends on how you value the risks in both deals.

In Finance, two essentials concepts:

- typically there is a compensation required for putting away money for a while $\rightarrow$ **nominal interest rates**
- typically investors require a compensation for the pain they experience when they lose money $\rightarrow$ **risk premia**

The **magnitude** of risk premia is determined by the investor's utility function: the utility function summarizes how an investor values the joy of gains versus the pain of losses.

OK, but what does this say about the "fair" price of the second investment opportunity?

The "fair" price of the second investment opportunity depends on **your risk aversion**, on the way you value a 100% loss versus an 11.1% gain.

If you are **risk-neutral**, meaning that the joy of 1 EUR gain equals the pain of 1 EUR loss, then

- you are indifferent between the two investment opportunities: all investments with **expected payoff** equal to EUR 1,000 are the same for you
- the required return on these investments is the nominal risk free rate
- the fair price is given by **discounting the expected payoff by the nominal risk free rate**

How does risk aversion change this game?

You would require a **compensation** for taking the credit risk on me.

There is just one mechanism through which the expected return goes up $\rightarrow$ **a lower price**.

The more risk averse you are, the less you are willing to pay for the second investment opportunity.

That brings me to the essential part of this clip:

In general it is incorrect to calculate the "fair" price of a future financial payoff by discounting the expected value of this financial payoff with the nominal risk free interest rate.

As we have seen, the **exception** is the case where the investor is risk-neutral: the investor requires the risk free return on all his investments then.

More formal: under risk aversion, the time t fair price of a payoff V at time $T \geq t$ is **not** given by the following general formula:

$$V_t = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{P}}(V_T),$$

where r is the continuously compounded nominal risk free interest rate that that is relevant for the period $[t, T]$ and $\mathbb{P}$ denotes the real-world probability measure.

Using this formula in our example would give a price of EUR 1,000, **no matter the level of risk aversion**.

Hence, you need to make a correction for risk aversion in the general formula in order to arrive at a different price.

And finally: what is the link with **no-arbitrage**?

Two rules that we know of at this stage with respect to no-arbitrage:

- The no-arbitrage price should be at least the minimum payoff of the instrument that is priced
- The Law of One Price: if there is a replicating strategy, the no-arbitrage price of the financial instrument equals the price of the replicating portfolio

In our example we assumed there is no market where the credit risk on Mark-Jan is traded: to exclude arbitrage opportunities the price should be larger than zero.

Clip 4: Pricing of Equity Forwards

Mark-Jan Boes

September 2020

Up to now we have talked about some general principles regarding **"fair"** and **"correct"** pricing of a financial instrument.

In the presence of a market, we typically say that the fair price is the price that **rules out arbitrage opportunities**, i.e. the no-arbitrage price.

We have identified two approaches to determine the no-arbitrage price of an instrument:

- Replication method: based on the Law of One Price
- Pricing kernel method: discount future expected cash-flows with rates that contain the market's aggregate attitude towards risk

I will illustrate the replication method for pricing of an **equity forward contract**.

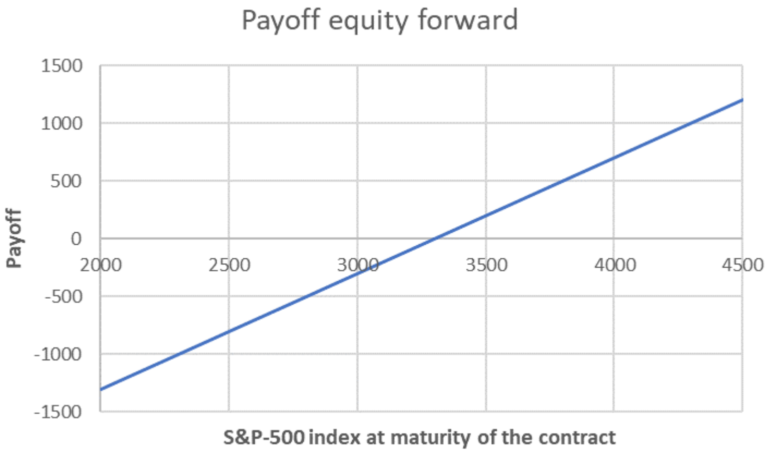
If you enter into a long position of an equity forward contract at time t that matures at time T with underlying value S and forward price K_t then you have the obligation to buy the underlying at at time T against price $K_{t,T}$.

At time t the two parties of the contract need to agree on the value for $K_{t,T}$, usually such that there is no exchange of cash at time t.

The payoff of the contract at the maturity date T is then given by:

$$V_T = S_T - K_{t,T}.$$

Graphically (taken from clip 2):



The pricing question we have over here is: what value should $K_{t,T}$ take in order to get $V_t = 0$?

Hence, we assume that the value of the contract at inception is known ($V_t = 0$) and we want to know which reference price $K_{t,T}$ fits, under no-arbitrage, with this value.

So we need to think of a strategy in a frictionless markets that has value zero at the start and delivers the payoff at the maturity of the equity forward contract.

In order to get replicate S_T perfectly, there is just one thing that we can do:

- buy the stock in the market against S_t

We need money to buy the stock, so we borrow cash amount S_t for a period $(T - t)$ against a continuously compounded interest rate r.

As result we have at portfolio value equal to zero at time t, exactly what we wanted.

At time T:

- The stock position has changed to value S_T
- The cash position has changed to value $-S_t e^{r(T-t)}$

So, at time T our replicating portfolio V takes value:

$$V_T = S_T - S_t e^{r(T-t)}.$$

This portfolio is constructed in such a way that it has value 0 at time t:

$$V_t = 0.$$

The Law of One Price says that two portfolios with identical payoffs (in each state of the world) should have the same price.

As said, the value for $K_{t,T}$ in the equity forward contract is determined such that the forward contract value equals 0 at time t: this is also the starting value of our replicating portfolio.

The only way to satisfy the Law of One Price is to choose $K_{t,T} = S_t e^{r(T-t)}$ in the equity forward contract.

Intuitively this makes sense: your counterparty passes the cost of delivering the stock (i.e. buying the stock 'plus' the funding cost) onto you.

Suppose that the reference price $K_{t,T}$ **is not equal to** $S_t e^{r(T-t)}$, how would the arbitrage work?

For example, $K_{t,T} > S_t e^{r(T-t)}$.

General rule: you buy the cheaper instrument and sell the more expensive instrument.

In this case:

- you "buy" the replicating strategy
- you "sell" (go short) the equity forward

It is obvious that this arbitrage portfolio A has value zero at inception, i.e. $A_t = 0$.

What do we have at the maturity date T of the equity forward?

$$\begin{aligned} A_T &= (S_T - S_t e^{r(T-t)}) - (S_T - K_{t,T}) \\ &= K_{t,T} - S_t e^{r(T-t)} > 0. \end{aligned}$$

So, you start with nothing and end up with something (guaranteed) $\implies$ **arbitrage opportunity**.

What about the **pricing kernel method**?

In Finance the following statements are equivalent (**Fundamental Theorem of Finance**):

- No arbitrage
- The existence of a positive **linear pricing rule**
- The existence of a finite positive optimal demand positive for some agent who prefers more to less

If you are interested in the full proof, you can look it up in Dybvig and Ross (1987).

In addition to the Fundamental Theorem of Finance, we have the **Representation Theorem** that says:

- The existence of a positive **linear pricing rule**
- The existence of a positive **pricing kernel (aka stochastic discount factor)**
- The existence of positive **risk-neutral probabilities** and an associated risk free rate

These theorems are at the heart of **Asset Pricing Theory**.

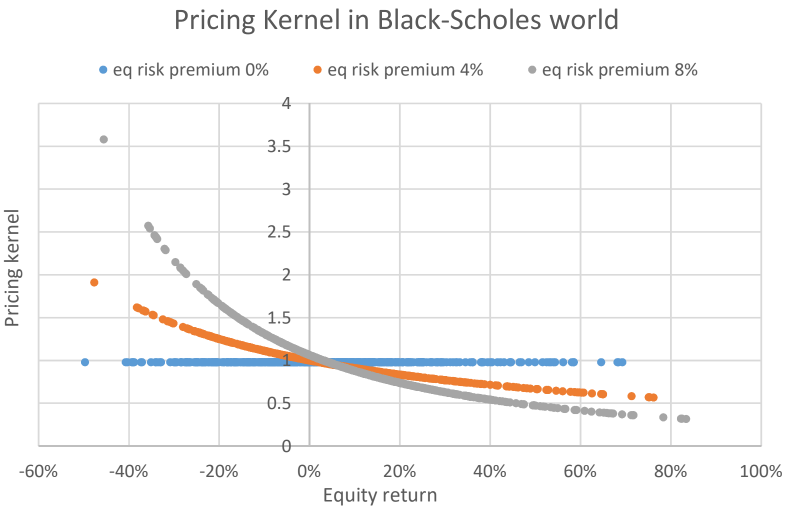
This basically means that the time t no arbitrage price of a future payoff V at time $T > t$ is given by:

$$V_t = \mathbb{E}_t^{\mathbb{P}}\left(V_T \frac{\pi_T}{\pi_t}\right),$$

where π represents the pricing kernel (aka the stochastic discount factor).

From an economic perspective, the pricing kernel summarizes the risk preferences of an investor: it defines the good states and bad states of the world.

The bad states of the world are those states where the marginal utility is high (a lot of joy from 1 Euro gain and a lot of pain of 1 Euro loss).


Pricing Kernel in Black-Scholes world
eq risk premium 0%
eq risk premium 4%
eq risk premium 8%
Pricing kernel
4
3.5
3
2.5
2
1.5
1
0.5
0
-60%
-40%
-20%
0%
20%
40%
60%
80%
100%
Equity return

Is there a clear intuition for the pricing kernel?

- Actually, that is not very easy
- A high value of the pricing kernel means that an investor appreciates a positive payoff (a lot) in that particular state of the world
- In the graph of the previous slide: we see for the gray line that the pricing kernel takes value 3 for an S&P-return of -40%, i.e. investors appreciates a financial instrument that provides positive payoffs if the S&P incurs big losses
- Translated into financial terminology: an investor is willing to pay a high price, i.e. accepts a negative expected return, for an instrument that provides a positive payoff when equity returns are strongly negative

- For instance, an instrument that pays 1 if the S&P loses 40% and 0 in all other states has an expected return equal to $\frac{1}{3} - 1$ $=$ -67% in case the gray line economy holds.
- An investor that can expect wealth losses if the S&P-500 loses a large amount will require a risk premium for taking on this investment: she wants to be compensated for the pain she experiences when such an event occurs
- The convexity of the line shows how risk averse investors are: the steeper the line, the more risk averse
- A flat line means that investors are risk-neutral: a 40% loss causes the same pain as the joy they experience from a 40% gain

If we want to find the no-arbitrage reference price for an equity forward using the pricing kernel method we need to solve the following equation:

$$0 = \mathbb{E}_t^{\mathbb{P}}\left((S_T - K_{t,T}) \frac{\pi_T}{\pi_t} \right),$$

where π represents the pricing kernel.

Purpose is to find the no-arbitrage value of $K_{t,T}$ such that no exchange of cash needed at inception of the contract.

We can rewrite the pricing equation for the forward contract as:

$$0 = \mathbb{E}_t^{\mathbb{P}}\left(S_T \frac{\pi_T}{\pi_t}\right) - K_{t,T}e^{-r(T-t)}.$$

Under no-arbitrage, the first part of the RHS is equal to:

$$\mathbb{E}_t^{\mathbb{P}}\left(S_T \frac{\pi_T}{\pi_t}\right) = S_t.$$

Using this we can derive the fair value of $K_{t,T}$:

$$K_{t,T} = S_t e^{r(T-t)}.$$

So, we had two pricing methods:

- Application of the Law of One price
- Pricing kernel method

Our conclusion: both pricing methods give exactly the same no-arbitrage reference price $K_{t,T}$.

This is very obvious: if the payoff can be **replicated perfectly** with base assets then the value of the replicating strategy should be the unique no-arbitrage value of the instrument.

Clip 5: Risk-neutral pricing

Mark-Jan Boes

September 2020

Up to now we have talked about some general principles regarding **"fair"** and **"correct"** pricing of a financial instrument.

In the presence of a market, we typically say that the fair price is the price that **rules out arbitrage opportunities**, i.e. the no-arbitrage price.

We have identified two approaches to determine the no-arbitrage price of an instrument:

- Replication method: based on the Law of One Price
- Pricing kernel method: discount future expected cash-flows with rates that contain the market's aggregate attitude towards risk

There is one other pricing method that is applied a lot in theory and practice: **the risk-neutral pricing method**.

Let me try to provide you the intuition of this method by means of an example.

Suppose you are a bookmaker:

- You are taking bets on a two-horse race
- You have studied the historical performance of both horses over various distances and have looked at factors such as training, diet and choice of jockey
- Using all this information you correctly calculate the probability of the first horse winning at 25% and, accordingly, a 75% probability that the second horse wins
- Consequently: the odds are set at **3-1 against** (horse 1) and **3-1 on** (horse 2)

What do these odds mean?

When a price is quoted in the form of n-m **against**, it means that a successful bet of \$$m$ will be rewarded with \$$n$ plus stake returned. The implied probability of victory is:

$$\frac{m}{m+n}.$$

In our example:

- **3-1 against** for horse 1: a \$1 bet will deliver \$4 if horse 1 wins
- **3-1 on** for horse 2: a \$1 bet will deliver \$1.333 if horse 2 wins

There is a degree of popular sentiment: the aggregate bet for horse 1 is $\$5,000$ and $\$10,000$ for horse 2.

Hence, the payoff for the bookmaker is:

- loss of $\$5,000$ if the first horse wins
- profit of $\$1,667$ if the second horse wins
- the expected profit therefore is $25\% \times -5,000 + 75\% \times 1,667 = 0$

It seems fine, but the bookmaker can suffer a large loss.

Suppose now that the odds were:

- **2-1 against** for horse 1: a \$1 bet will deliver \$3 if horse 1 wins
- **2-1 on** for horse 2: a \$1 bet will deliver \$1.5 if horse 2 wins

Now, the payoff of the bookmaker becomes (we assume that the bets on both horse don't change):

- profit of \$0 if the first horse wins
- profit of \$0 if the second horse wins
- the expected profit therefore is $25\% \times 0 + 75\% \times 0 = 0$

The outcome of the horse race is **irrelevant** for the bookmaker!

What the example shows:

- **Two possible outcomes** of a random experiment: horse 1 wins or horse 2 wins
- Probabilities of these outcomes occurring: 25% horse 1 and 75% horse 2
- We call these **"real-world probabilities"** as these are derived from real-world information
- For pricing a different set of probabilities were used: the bookmaker calculates the odds such that there is no risk for him
- These probabilities are called **risk-neutral probabilities**: 33% of horse 1 winning and 67% of horse 2 winning

Later on this clip and this course we will see that this concept is relevant for no-arbitrage pricing as well.

In the previous clip we claimed that in Finance the following statements are equivalent:

- No arbitrage
- The existence of a positive **linear pricing rule**
- The existence of a finite positive optimal demand positive for some agent who prefers more to less

If you are interested in the full proof, you can look it up in Dybvig and Ross (1987).

In addition to the Fundamental Theorem of Finance, we have the **Representation Theorem** that says:

- The existence of a positive **linear pricing rule**
- The existence of a positive **pricing kernel (aka stochastic discount factor)**
- The existence of positive **risk-neutral probabilities** and an associated risk free rate

These theorems are at the heart of **Asset Pricing Theory**.

Suppose we have a state space, Ω, with a **finite number** of m outcomes

$$\begin{bmatrix}\theta_1 & \theta_2 & \cdots & \theta_m\end{bmatrix}$$

The model is in **discrete time**: specifically, we have time $t = 0$ and $t = 1$. At time $t = 1$ all uncertainty is resolved and the world is in one and only one of the m states of nature in Ω.

We will also assume that there are a finite number, n, of traded assets with current price vector ν:

$$\begin{bmatrix}\nu_1 & \nu_2 & \cdots & \nu_n\end{bmatrix}$$

Next, we have an $m \times n$ **payoff matrix**, G:

$$\begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{m1} & \cdots & g_{mn} \end{bmatrix}$$

Each row represents lists the payoffs of the n securities in that particular state of nature.

The Fundamental Theorem states that no-arbitrage is equivalent to the existence of a positive linear pricing rule.

In our model this implies the existence of a positive $1 \times m$ vector η such that:

$$\nu = \eta G.$$

The vector η maps the payoffs in matrix G to no-arbitrage prices in vector ν.

There is a close relation between the **linear pricing vector** η, **the pricing kernel** and **risk-neutral probabilities**.

We define the pricing kernel as a $1 \times m$ vector π such that the value V of any payoff z ($m \times 1$ vector) is given by:

$$V(z) = \mathbb{E}^{\mathbb{P}}(\pi z) = \mathbb{E}^{\mathbb{P}}(\pi)\mathbb{E}^{\mathbb{P}}(z) + \mathbb{C}\mathrm{ov}(\pi, z).$$

Using the definition of expectation we can rewrite:

$$V(z) = \mathbb{E}^{\mathbb{P}}(\pi z) = \sum_{i=1}^{m} p_i \pi_i z_i,$$

where p_i is the probability of state of nature i realizing at time 1.

How can we relate the linear pricing verctor η to the pricing kernel π?

Let us look at the **price** of a state versus the **probability** of a state:

$$\pi_i = \frac{\eta_i}{p_i} > 0.$$

Then,

$$V(z) = \sum_{i=1}^{m} \eta_i z_i = \sum_{i=1}^{m} p_i \frac{\eta_i}{p_i} z_i = \sum_{i=1}^{m} p_i \pi_i z_i = \mathbb{E}^{\mathbb{P}}(\pi z).$$

Hence, a positive linear pricing rule implies the existence of a positive pricing kernel.

Consequently, if we have a model for π we can calculate the no-arbitrage price of a payoff by means of calculating a real-world expectation.

And how about **risk-neutral probabilities**?

We could define risk-neutral probabilities as:

$$q_i = \frac{\eta_i}{\sum_{i=1}^m \eta_i}.$$

We have,

$$V(z) = \sum_{i=1}^m \eta_i z_i = \left(\sum_{i=1}^m \eta_i\right) \sum_{i=1}^m \frac{\eta_i}{\left(\sum_{i=1}^m \eta_i\right)} z_i = \left(\sum_{i=1}^m \eta_i\right) \sum_{i=1}^m q_i z_i$$
$$= \left(\sum_{i=1}^m \eta_i\right) \mathbb{E}^{\mathbb{Q}}(z).$$

Why do we call these new probabilities **risk-neutral probabilities**?

Suppose we have an asset with the following payoff structure:

$$\iota = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

By no-arbitrage the price of this payoff is given by:

$$V(\iota) = \eta\iota = \sum_{i=1}^{m} \eta_i.$$

However, in the presence of a risk free asset paying a rate r between times 0 and 1, we also know (Law of One Price) that:

$$V(\iota) = \frac{1}{1+r}.$$

Therefore, we must have:

$$\sum_{i=1}^{m} \eta_i = \frac{1}{1+r}.$$

Now returning to the earlier risk-neutral valuation formula:

$$V(z) = \left(\sum_{i=1}^{m} \eta_i\right) \mathbb{E}^{\mathbb{Q}}(z) = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}(z).$$

Hence, the expected return on an arbitrary payoff vector z is given by:

$$\mathbb{E}^{\mathbb{Q}}(R^z) = \frac{\mathbb{E}^{\mathbb{Q}}(z)}{V(z)} = (1+r).$$

Under the new probabilities q **all assets earn the risk free rate (in expectation)**, independent of their riskiness $\rightarrow$ **risk-neutral world**.

Basically, we have introduced an alternative way of calculating the no-arbitrage price of a derivative: **the risk-neutral pricing method**.

The idea:

- There is an outcome space, i.e. possible outcomes of a random experiment in **the real-world**, e.g. horse 1 wins or horse 2 wins
- There are **real-world probabilities** associated with the outcomes of the experiment, e.g. a 25% probability that horse 1 wins
- If we want to use these probabilities for pricing, we need to take into account risk preferences (see previous clips) $\rightarrow$ **pricing kernel method**

The idea (cont'd):

- Another way is to adjust the probabilities such that all assets earn the risk-free rate in expectation
- If we have these adjusted, so-called risk-neutral, probabilities then pricing comes down to evaluating an expectation using these probabilities and discount this expectation with the risk free rate (see previous slides)

The **Fundamental Theorem of Finance** and the **Representation Theorem** ensure that the resulting pricing is a **no-arbitrage price**.

All the previous connects the more general asset pricing literature and the more specific (and mathematical) derivatives pricing literature.

What we are going to do in this course:

- Define an outcome space
- Define **real-world** probabilities on this outcome space
- Apply a trick to **move** from the **real-world probabilities** to the **risk-neutral probabilities**
- Find the **no-arbitrage value** by evaluation the risk-neutral expectation of the payoff and discount this with the risk free rate

Where possible, we will apply both the replication method and the risk-neutral pricing method, for instance, in the setting of binomial trees (next clip).

Clip 6: One-period binomial trees

Mark-Jan Boes

September, 2020

The **purpose** of this clip is to find the **no-arbitrage price** of a payoff in **the binominial tree model**.

For pricing we will use:

- The replication method
- The risk-neutral valuation method

Where are **stochastic processes** in this picture?

- In order to replicate a derivative in a general case, we need to understand how the price of the underlying asset behaves in time.
- We model it by a random or stochastic process.
- Additionally, we need to understand how to model the replicating portfolio and the derivative, when the underlying process is stochastic.

The binomial tree model assumes a simple stochastic process for the underlying asset of the derivative:

- Two assets: the underlying value of the derivative and a bond
- **Discrete time model**: only two time points
- **Discrete probability distribution** of the underlying asset: only two possible outcomes
- Bond price process is deterministic

Graphically:

$$\begin{tikzcd}[column sep=huge]
 & S_1(H) = u \cdot S_0 \\
S_0 \arrow[ur, "p"] \arrow[dr, "(1-p)"'] & \\
 & S_1(T) = d \cdot S_0
\end{tikzcd}$$

$$\begin{tikzcd}[column sep=huge]
B_0 \arrow[r] & B_1 = (1+r) \cdot B_0
\end{tikzcd}$$

$t = 0$ $\qquad\qquad\qquad$ $t = 1$

We assume:

- $u > d$
- $S_0, S_1 > 0$

Let us first **rule out arbitrage opportunities** between the underlying value S and the bond B.

The binomial model is *free of arbitrage* if and only if the following relations hold:

$$d \leq 1 + r \leq u.$$

The intuition is simple: one of the outcomes for S at time $t = 1$ should be lower or equal than B otherwise everyone would sell the bond and buy the stock.

Condition $d \leq 1 + r$: Let's **assume the opposite** $u > d > 1 + r$.

- We borrow money (short x bonds) and buy y shares such that

$$x \cdot B_0 = y \cdot S_0.$$

  This is a *zero-investment strategy*, where we are short bond and long stock. The initial value of our portfolio is $V_0 = 0$.

- Even in the worst case we have enough money to repay our debt and have some money left.

$$y \cdot S_1(T) = y \cdot d \cdot S_0 = d \cdot x \cdot B_0 > x \cdot (1 + r) B_0$$

  We have *zero risk*: $\mathbb{P}[V_1^h > 0] = 1$.

- Our assumption that $d > 1 + r$ **leads to an arbitrage!**

We can also assume that $d < u < 1 + r$:

- We then use the opposite strategy to short the stock and invest in the bond. This will again lead to the arbitrage.
- Thus, we proved the *if* statement. We proved that *if there is no arbitrage in the market*, we must have $d \leq 1 + r \leq u$.
- It is also possible to prove the *only-if* statement: If the relations $d \leq 1 + r \leq u$ hold, then it is not possible to create a portfolio such that the initial value of the portfolio is zero

$$V_0 = yS_0 - xB_0 = 0,$$

and the probability of nonpositive values of the portfolio at time one

$$V_1 = yS_1 - x(1 + r)B_0$$

is zero. This will prove the absence of arbitrage.

We want to use the one-period binomial model to price a derivative security that pays $V_1(H)$ and $V_1(T)$, for a head and tail, respectively.

We do this under the assumption of availability of two base instruments (stock and bond) in an arbitrage free binomial tree model.

In our notation we already take some aspects from **probability theory** over here:

- heads and tails are the two possible outcomes of a natural experiment: tossing a coin
- the outcomes occur with a certain probability: p for heads and $(1-p)$ for tails in this case
- the numerical realisation of random variable S at time 1 depends on the outcome of the natural experiment: if heads comes up then S takes value $S_1 = u \cdot S_0$.

Graphically:

$$
\begin{tikzcd}
 & V_1(u \cdot S_0) = V(H) \\
V_0(S_0) = ? \arrow[ur, "p"] \arrow[dr, "(1-p)"'] & \\
 & V_1(d \cdot S_0) = V(T)
\end{tikzcd}
$$

$$B_0 \longrightarrow B_1 = (1+r) \cdot B_0$$

$$t = 0 \qquad\qquad t = 1$$

From previous clips we know that the **wrong way** of calculating V_0 is:

$$V_0 = \frac{1}{1+r}\left[p \cdot V(H) + (1-p) \cdot V(T)\right] = \frac{1}{1+r}\mathbb{E}_0^{\mathbb{P}}\left[V_1\right].$$

If this were the correct price, the investor in the derivative instrument accepts an expected rate of r, which is the rate of return on the riskless bond B.

It could be, however, that the investor is risk averse and wants to be compensated for the risks in the contract or wants to pay more if the derivative pays off in the bad states of the world, i.e. states where the pricing kernel is high.

Let us try the **replication method**:

- Suppose we have **initial wealth** X_0 and buy Δ_0 shares of stock at time zero.
- This leaves us with the **bond position** $X_0 - \Delta_0 S_0$, could be either positive or negative.
- At time one, the value of our portfolio is

$$X_1 = \Delta_0 S_1 + (1+r)(X_0 - \Delta_0 S_0) = (1+r)X_0 + \Delta_0(S_1 - (1+r)S_0)$$

Note that X_1 is random. It depends on the outcome of the coin toss.

S_0

Δ_0

$X_0 = \Delta_0 S_0 + (X_0 - \Delta_0 S_0)$

$\xrightarrow{p}$

$S_1(H) = u \cdot S_0$

$X_1(H) = \Delta_0 S_1(H) + (1+r)(X_0 - \Delta_0 S_0) = V_1(H)$

$\xrightarrow{(1-p)}$

$S_1(T) = d \cdot S_0$

$X_1(T) = \Delta_0 S_1(T) + (1+r)(X_0 - \Delta_0 S_0) = V_1(T)$

In order to apply the Law of One Price, we need to choose X_0 and Δ_0 so that $X_1(H) = V_1(H)$ and $X_1(T) = V_1(T)$.

X_0 will then be the no-arbitrage price of the derivative contract.

We have a system of two equations of two variables X_0 and Δ_0

$$\begin{cases} \Delta_0 S_1(T) + (1+r)(X_0 - \Delta_0 S_0) = V_1(T) \\ \Delta_0 S_1(H) + (1+r)(X_0 - \Delta_0 S_0) = V_1(H). \end{cases}$$

The solution for Δ_0 follows easily:

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}, \quad \text{or} \quad \Delta_0 = \frac{V_1(H) - V_1(T)}{(u-d) \cdot S_0},$$

Using the solution for Δ_0, we solve for X_0:

$$\begin{aligned}(1+r)X_0 &= V_1(T) - \Delta_0 S_1(T) + (1+r)\Delta_0 S_0 \\ &= V_1(T) - \frac{V_1(H) - V_1(T)}{(u-d)\cdot S_0} dS_0 + (1+r)\frac{V_1(H) - V_1(T)}{(u-d)S_0} S_0 \\ &= V_1(H)\cdot\left(\frac{1+r-d}{u-d}\right) + V_1(T)\cdot\left(\frac{u-1-r}{u-d}\right).\end{aligned}$$

Hence, we can easily calculate the no-arbitrage price using the information that we have available at time $t = 0$.

We can be **sure** that it is a no-arbitrage price because we designed at strategy of base instruments (through Δ_0 and using X_0) that exactly delivers the payoff of the derivatives in all states (two in this case) of the world $\implies V_0 = X_0$.

From the formula on the previous slide we can immediately see the connection with the **risk-neutral pricing method**:

$$X_0 = \frac{1}{1+r}\left[\tilde{p}V_1(H) + (1-\tilde{p})V_1(T)\right],$$

where

$$\tilde{p} = q = \frac{1+r-d}{u-d}, \qquad 1-\tilde{p} = 1-q = \frac{u-1-r}{u-d}.$$

Due to the no-arbitrage condition $\tilde{p} > 0$ and less than one (also holds for $(1-\tilde{p})$ then) $\implies$ we have defined a **new probability measure**.

The new probabilities are also called **risk-neutral probabilities**. Why is that?

That is because in a world where these new probabilities hold, all assets earn the risk free rate (in expectation).

We are able to verify this in the setting above for the stock price S:

$$\begin{aligned}
\mathbb{E}_0^{\mathbb{Q}}(S_1) &= \left(\frac{1+r-d}{u-d}\right) S_0 u + \left(\frac{u-1-r}{u-d}\right) S_0 d \\
&= S_0 \left(\frac{(1+r-d)u + (u-1-r)d}{u-d}\right) \\
&= S_0 \left(\frac{(1+r)u - (1+r)d}{u-d}\right) \\
&= S_0 \left(\frac{(1+r)(u-d)}{u-d}\right) = S_0(1+r).
\end{aligned}$$

The payoff of an equity forward with maturity date $t = 1$ and reference price K is given by:

$$V_1 = S_1 - K.$$

In our model, Δ_0 is given by:

$$\Delta_0 = \frac{(S_0 u - K) - (S_0 d - K)}{S_0 u - S_0 d} = 1.$$

We knew this already from earlier clips: if you want to replicate the payoff of the obligation of buying the stock in the future, you buy the stock right now.

Which reference price K leads to a no-arbitrage value of 0?

We use $\Delta_0 = 1$ and $X_0 = 0$:

$$(1+r) \cdot 0 = (S_0 d - K) - 1 \cdot S_0 d + (1+r) \cdot 1 \cdot S_0.$$

This implies:

$$K = S_0 \cdot (1+r).$$

This is the same result as in earlier clips (now with a discretely compounded interest rate).

Clip 7: Multiperiod binomial trees

Mark-Jan Boes

September, 2020

The **purpose** of this clip is to find the **no-arbitrage price** of a payoff in **the multiperiod binominial tree model**.

For pricing we will use:

- The replication method
- The risk-neutral valuation method

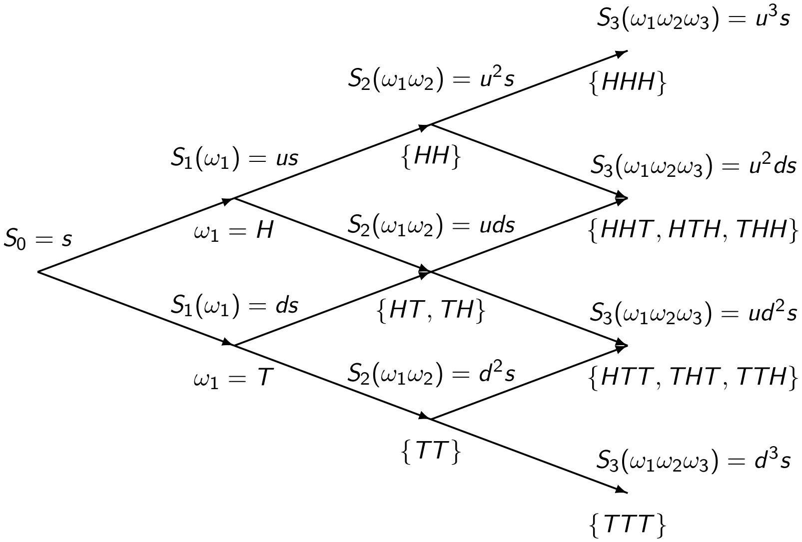
In the previous clip we used the **one-period** binomial model as our model for the “world”.

The binomial tree model assumes a simple stochastic process for the underlying asset of the derivative:

- Two assets: the underlying value of the derivative and a bond
- **Discrete time model**: only two time points
- **Discrete probability distribution** of the underlying asset: only two possible outcomes
- Bond price process is deterministic

In the **the multiperiod binominial tree model** we take more than two time points. In addition, from each node two possible outcomes for the underlying are possible.

Example: **3-period binomial model**.



Let us try to find the no-arbitrage price of a payoff V that matures at time $t = 3$ by means of the **replication method**.

We use the same approach as in the one-period binomial tree model:

- Suppose we have **initial wealth** X_0 and buy Δ_0 shares of stock at time zero.
- This leaves us with the **bond position** $X_0 - \Delta_0 S_0$, could be either positive or negative.

At time one, the value of our portfolio is

$$X_1 = \Delta_0 S_1 + (1+r)(X_0 - \Delta_0 S_0) = (1+r)X_0 + \Delta_0(S_1 - (1+r)S_0)$$

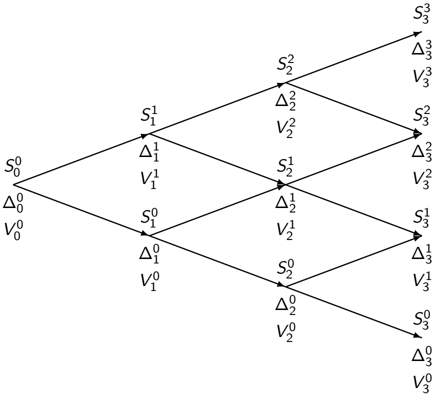
Note that X_1 is random. It depends on the outcome of the coin toss.

With the portfolio value that we have available at time $t = 1$, which depends on the node, we are going to do the same thing as at time $t = 0$: we buy Δ_1^i shares and invest the remainder in the bond.

This leads to the following possible values of X at time $t = 2$:

$$
\begin{aligned}
X_2^0 &= \Delta_1^0 S_2^0 + (1+r)(X_1^0 - \Delta_1^0 S_1^0) \\
X_2^1 &= \Delta_1^1 S_2^1 + (1+r)(X_1^1 - \Delta_1^1 S_1^1) \\
X_2^2 &= \Delta_1^1 S_2^2 + (1+r)(X_1^1 - \Delta_1^1 S_1^1)
\end{aligned}
$$

And at time $t = 3$, portfolio value X can take four different values.



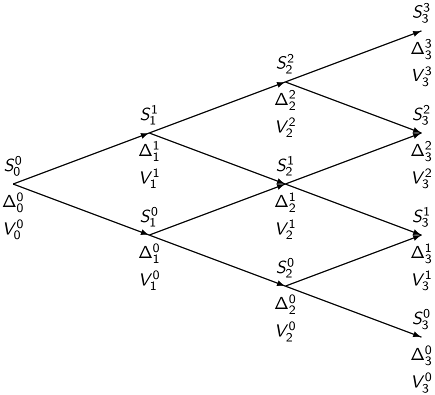
$$X_1^1 = \Delta_0^0 S_1^1 + (1+r)(X_0^0 - \Delta_0^0 S_0^0)$$
$$X_1^0 = \Delta_0^0 S_1^0 + (1+r)(X_0^0 - \Delta_0^0 S_0^0)$$
$$\vdots$$
$$X_2^2 = \Delta_1^1 S_2^2 + (1+r)(X_1^1 - \Delta_1^1 S_1^1)$$
$$\vdots$$
$$X_3^1 = \Delta_2^0 S_3^1 + (1+r)(X_2^0 - \Delta_2^0 S_2^0)$$
$$X_3^0 = \Delta_2^0 S_3^0 + (1+r)(X_2^0 - \Delta_2^0 S_2^0)$$

We apply the same principle as for the one-period binomial tree: choose X_0^0 and the Δs in such a way that the payoff of the strategy is exactly the same as the payoff of the derivative.

Idea of the approach:

- We have knowledge of the derivative value at time $t = 3$ for each possible outcome of S_3
- With this knowledge we are able to solve three binomial one-period problems at time $t = 2$
- This leads to three values of the derivatives at time $t = 2$: V_2^0, V_2^1 and V_2^2
- Then we can repeat the same exercise for time $t = 1$ and $t = 0$, ultimately leading to the value of V_0^0

In each node we can apply **the same formulas** as in the one-period model:



$$\Delta_2^0 = \frac{V_3^1 - V_3^0}{S_3^1 - S_3^0}$$

$$\Delta_2^1 = \frac{V_3^2 - V_3^1}{S_3^2 - S_3^1}$$

$$\Delta_2^2 = \frac{V_3^3 - V_3^2}{S_3^3 - S_3^2}$$

The Δs can be used to find the no-arbitrage values V_2^0, V_2^1 and V_2^2 in nodes S_2^0, S_2^1 and S_2^2, respectively.

For instance, we know from the previous clip that we can calculate V_2^2 as:

$$V_2^2 = \frac{1}{1+r}\left(V_3^2 - \Delta_2^2 S_3^2 + (1+r)\Delta_2^2 S_2^2\right).$$

We can apply the **risk-neutral valuation method** in the same spirit as in the previous clip.

The formula on the previous slide can be written in such a way that it becomes independent of Δ_2^2:

$$\begin{aligned} V_2^2 &= \frac{1}{1+r}\left(V_3^2 - \Delta_2^2 S_3^2 + (1+r)\Delta_2^2 S_2^2\right) \\ &= \frac{1}{1+r}\left(V_3^3 \cdot \left(\frac{1+r-d}{u-d}\right) + V_3^2 \cdot \left(\frac{u-1-r}{u-d}\right)\right) \\ &= \frac{1}{1+r}\left(q \cdot V_3^3 + (1-q) \cdot V_3^2\right). \end{aligned}$$

Hence, we can calculate V_2^2 by discounting the expected value of V_3 **conditional on** S_2^2 with the risk free rate r, where the expectation is calculated using **the risk-neutral probabilities**.

Clip 1: An Introduction into Probability Theory

Mark-Jan Boes

September, 2020

In this clip we will treat the following concepts in the context of the multiperiod binomial tree:

- Probability space: Ω
- Events: A
- Algebra: $\mathcal{F}$
- Filtration: $\mathcal{F}_t$
- Adapted to filtration
- Probability measure: $\mathbb{P}$
- Random variable: X
- Expectation: $\mathbb{E}$
- Conditional expectation: $\mathbb{E}_t$
- Martingale: $\mathcal{M}$
- Markov process

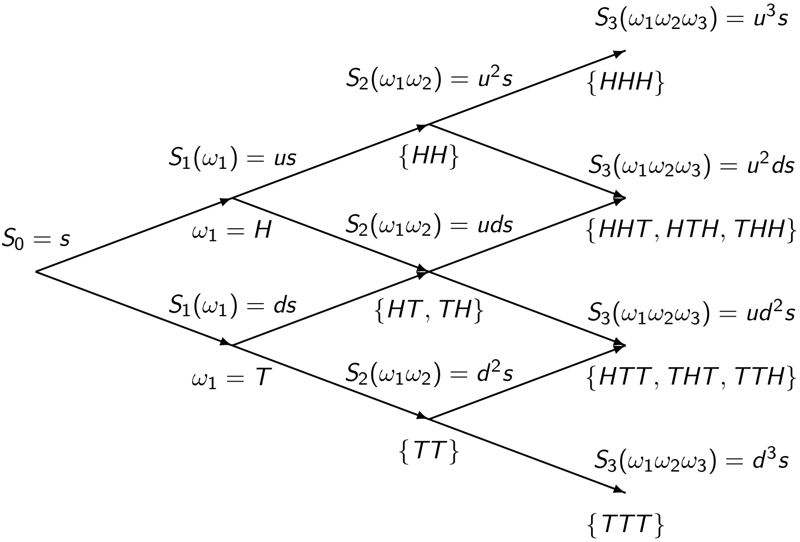
Our general purpose of this week (and next week) is that we want to get an understanding of the following:

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sigma-algebra such that $\mathcal{G} \subseteq \mathcal{F}$, and let X be a random variable that is either nonnegative or integrable.

The **conditional expectation** of X given $\mathcal{G}$, denoted $\mathbb{E}[X|\mathcal{G}]$, is any random variable that satisfies:

1. $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable.
2. For every $B \in \mathcal{G}$ it holds that

$$\int_B \mathbb{E}[X|\mathcal{G}](\omega) d\mathbb{P}(\omega) = \int_B X(\omega) d\mathbb{P}(\omega),$$


$S_3(\omega_1\omega_2\omega_3) = u^3s$
$\{HHH\}$
$S_2(\omega_1\omega_2) = u^2s$
$\{HH\}$
$S_3(\omega_1\omega_2\omega_3) = u^2ds$
$S_1(\omega_1) = us$
$\omega_1 = H$
$S_0 = s$
$S_2(\omega_1\omega_2) = uds$
$\{HHT, HTH, THH\}$
$S_1(\omega_1) = ds$
$\{HT, TH\}$
$S_3(\omega_1\omega_2\omega_3) = ud^2s$
$\omega_1 = T$
$S_2(\omega_1\omega_2) = d^2s$
$\{HTT, THT, TTH\}$
$\{TT\}$
$S_3(\omega_1\omega_2\omega_3) = d^3s$
$\{TTT\}$

Let us look at finite probability spaces:

- Suppose we toss a coin three times. Our *sample space* Ω is

$$\Omega = \{\omega_1\omega_2\omega_3\} = \\ = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

- The subsets of Ω are called *events*.
- For example, the event "The first toss is a head" A_1 is

$$A_1 = \{\omega \in \Omega; \omega_1 = H\} = \{HHH, HHT, HTH, HTT\}.$$

- For example, the event "The first toss is a tail" A_2 is

$$A_2 = \{\omega \in \Omega; \omega_1 = T\} = \{THH, THT, TTH, TTT\}.$$

- If we take a union of A_1 and A_2 we get our sample space $\Omega = A_1 \cup A_2$.
- If we take a complement of A_1, i.e. all elements in Ω that are not in A_1 we get $A_1^c = A_2$. Similar, $A_2^c = A_1$.

There is more to say:

- The whole sample space Ω is also an event, which is called the *sure event* since one of its outcomes must always occur.
- The complement of Ω is the empty set $\emptyset$, which is also defined as an event but never occurs.

Our purpose is to summarize the essential probabilistic information that characterizes an experiment.

In probability theory this information is summarized in the triplet $(\Omega, \mathcal{F}, \mathbb{P})$, where

- Ω is set of possible outcomes, **the sample space**
- $\mathcal{F}$ is a **collection of events** that satisfies certain properties (next slide)
- $\mathbb{P}$ denotes the **probabilities** that are defined on these events

Definition

A family $\mathcal{F}$ of subsets of Ω is an **algebra** if the following holds:

1. $\emptyset \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. $A_i \in \mathcal{F} \quad \text{for} \quad i = \overline{1, N} \quad \Rightarrow \quad \bigcup_i^N A_i \in \mathcal{F}$

On the previous slide we introduced the triplet $(\Omega, \mathcal{F}, \mathbb{P})$: $\mathcal{F}$ is an algebra.

Consider again the multiperiod binomial tree.

- How can we make algebras in this model? We look at the paths that are possible at a particular point in time.
- At time t_0, all paths are possible. The collection of possible paths is $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
- At time t_1, the collection of possible paths depends upon the position

$$A_1 = \{HHH, HHT, HTH, HTT\} \quad \text{if } S_1(H)$$
$$A_2 = \{THH, THT, TTH, TTT\} \quad \text{if } S_1(T)$$

  Let's complete this collection with all possible unions and complements and we get the collection (algebra) $\mathcal{F}_1 = \{\emptyset, \Omega, A_1, A_2\}$.
- Note that $\mathcal{F}_0 \subseteq \mathcal{F}_1$.

- At time t_2:

$$B_1 = \{HHH, HHT\} \quad \text{if } S_2(HH), \quad B_2 = \{HTH, HTT\} \quad \text{if } S_2(HT),$$
$$B_3 = \{THH, THT\} \quad \text{if } S_2(TH), \quad B_4 = \{TTH, TTT\} \quad \text{if } S_2(TT).$$

- With unions and complements, we have

$$\begin{aligned}\mathcal{F}_2 = \{&\emptyset, \Omega, A_1(B_1 \cup B_2), A_2(B_3 \cup B_4), B_1, B_2, B_3, B_4,\\ &\text{and all unions of } B_i:\\ &B_1 \cup B_3, B_1 \cup B_4, B_2 \cup B_3, B_2 \cup B_4\\ &B_1 \cup A_2, B_2 \cup A_2, B_3 \cup A_1, B_4 \cup A_1\}.\end{aligned}$$

- Consider for example subset $B_1 \cup B_3$ ("the second toss is a head"). At time t_1 we could not say if this event happened or not. At time t_2 we can. Consider this as a "flow of information".
- Again, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2$.

- At time t_3, we know exactly which path was taken. Thus, the events in this case are all individual paths, e.g $C_1 = \{HHH\}$ and $C_3 = \{HTH\}$. The total collection with all unions and complements is

$$\mathcal{F}_3 = \{\emptyset, \Omega, A_1, A_2, B_1(C_1 \cup C_2), \dots, B_4(C_7 \cup C_8), C_1, \dots, C_8, \\ \text{and all remaining unions of } C_i\}$$

and again

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3.$$

- We showed that $\mathcal{F}_i$ are algebras by construction. An increasing set of algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3$ is called **filtration**.
- We can consider this as a *flow of information*, i.e. what is known until a particular time point.
- If a process X_n depends only on n first tosses, we say it is **adapted to filtration** $\mathcal{F}_n$. Given the information up to time t_n, we can say exactly what happened with X_n.

Now we look at probabilities:

- Consider again our *sample space* Ω

$$\begin{aligned}\Omega &= \{\omega_1\omega_2\omega_3\} = \\ &= \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.\end{aligned}$$

- If the probability of a head is p and a tail is q, then the probability of single element is

$$\mathbb{P}(\omega = \omega_1\omega_2\omega_3) = p^{\#H}q^{\#T}.$$

- For the event "The first toss is a head"

$$A_1 = \{\omega \in \Omega; \omega_1 = H\} = \{HHH, HHT, HTH, HTT\}.$$

- The probability of the event is the *sum of probabilities* of the elements

$$\begin{aligned}\mathbb{P}(A_1) &= \mathbb{P}(HHH) + \mathbb{P}(HHT) + \mathbb{P}(HTH) + \mathbb{P}(HTT) \\ &= (p^3 + p^2q) + (p^2q + pq^2) = p^2(p+q) + pq(p+q) \\ &= p^2 + pq = p(p+q) = p\end{aligned}$$

A **finite probability space** consists of three elements:

- Ω, a nonempty set, called the sample space which contains all possible outcomes of a random experiment (in our case: tossing a coin three times)
- $\mathcal{F}$, an algebra of subsets of Ω
- $\mathbb{P}$, a probability measure on $(\Omega, \mathcal{F})$, i.e. $\mathbb{P}$ is function that assigns to each element $\omega \in \Omega$ (or to each set $A \in \mathcal{F}$) a number in $[0, 1]$ such that
$$\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1.$$

An event is a subset of Ω and the probability of an event A is

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega).$$

Note that it follows from our definition that if subset A and subset B do not have common elements, i.e. $A \cap B = \emptyset$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a finite probability space. A *random variable* is a real-valued function defined on Ω.

For the three-period binomial tree we have 4 random variables:

$$S_0(\omega_1\omega_2\omega_3) = s \quad \forall \quad \omega_1\omega_2\omega_3 \in \Omega$$

$$S_1(\omega_1\omega_2\omega_3) = \begin{cases} u \cdot s & \text{if } \omega_1 = H \\ d \cdot s & \text{if } \omega_1 = T \end{cases}$$

$$S_2(\omega_1\omega_2\omega_3) = \begin{cases} u^2 \cdot s & \text{if } \omega_1 = \omega_2 = H \\ ud \cdot s & \text{if } \omega_1 \neq \omega_2 \\ d^2 \cdot s & \text{if } \omega_1 = \omega_2 = T \end{cases}$$

$$S_3(\omega_1\omega_2\omega_3) = \begin{cases} u^3 \cdot s & \text{if } \omega_1 = \omega_2 = \omega_3 = H \\ u^2 d \cdot s & \text{if there are two heads and one tails} \\ ud^2 \cdot s & \text{if there is one tail and two heads} \\ d^3 \cdot s & \text{if } \omega_1 = \omega_2 = \omega_3 = T. \end{cases}$$

Or we can consider S_i as a **process**, where i is time. In this way, we introduced a **discrete time stochastic process** with **discrete values**.

- We introduced probability $\mathbb{P}(\omega)$ for every element ω of space Ω and we know how to compute probability of a particular event A.
- We can define an event as the subset of outcomes in Ω for which the random variable equal to a particular value. In this way we introduce the *probability distribution* of the random variable as the probability of the random variable to take particular values.
- For example for binomial tree, the probability to be at node $k = 0, \ldots, n$ after n steps is

$$\mathbb{P}\binom{n}{k} = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

- It is easy to see that $\sum_{k=0}^{n} \mathbb{P}\binom{n}{k} = 1$.

Definition

Let X be a random variable defined on a finite probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *expectation* (or expected value) of X is defined to be

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega).$$

The variance of X is

$$Var[X] = \mathbb{E}\left[X - \mathbb{E}[X]\right]^2$$

From the definition it follows *linearity* of expectation

$$\mathbb{E}[c_1 X + c_2 Y] = c_1\mathbb{E}[X] + c_2\mathbb{E}[Y].$$

- Consider our binomial tree. We can compute the expectation of stock price at every time.
- But we can also ask what is the expectation of the stock price at time n given the value of the stock price at time $m < n$?
- We call this value the *conditional expectation* of S_n and denote as $E_m[S_n]$ or $E[S_n | \omega_1 \dots \omega_m]$.
- At time zero we do not know the value S_m, so conditional expectation is a *random variable* itself.

Definition

Suppose path $\omega_1 \dots \omega_m$ is given, we define

$$\begin{aligned} E_m[X_n] &= E[X_n | \omega_1 \dots \omega_m] \\ &= \sum_{\omega_{m+1} \dots \omega_n} p^{\#H(\omega_{m+1} \dots \omega_n)} q^{\#T(\omega_{m+1} \dots \omega_n)} X(\omega_1 \dots \omega_m \omega_{m+1} \dots \omega_n) \end{aligned}$$

$$E_0[X_n] = E[X_n], \quad E_n[X_n] = X_n.$$

Properties

1. **Linearity of conditional expectation**

$$E_m[c_1 X + c_2 Y] = c_1 E_m[X] + c_2 E_m[Y].$$

2. **“Take out what is known”**

$$E_m[XY] = X E_m[Y], \quad \text{if } X = X_m.$$

3. **Iterated conditioning** or **“Tower” property**

$$E_m[E_n[Y]] = E_m[Y], \quad \text{if } m \leq n.$$

4. **Independence**

$$E_m[X] = E[X], \text{if } X = X(\omega_{m+1} \ldots \omega_N).$$

1. Follows from the linearity of expectation.
2. X_m is know at time t_m and can be taken out of expectation as being non-random.
3. This is a version of the law of total probability

$$E_m[E_n[Y_N]] = \sum_n \left(\mathbb{P}(m,n)E_n[Y_N]\right) = \sum_n \left(\mathbb{P}(m,n)\sum_N \mathbb{P}(n,N)Y_N\right)$$

$$= \sum_N Y_N \left(\sum_n \mathbb{P}(m,n)\mathbb{P}(n,N)\right) = \sum_N Y_N \mathbb{P}(m,N) = E_m[Y_N]$$

4. Consequence of X being not depend on the path until time t_m

$$E_m[X] = \sum_{\omega_1\ldots\omega_m} \mathbb{P}(\omega_1 \ldots \omega_m) \cdot \sum_{\omega_{m+1}\ldots\omega_N} \mathbb{P}(\omega_{m+1}\ldots\omega_N)X(\omega_{m+1}\ldots\omega_N)$$

$$= \sum_{\omega_1\ldots\omega_N} \mathbb{P}(\omega_1\ldots\omega_N)X(\omega_{m+1}\ldots\omega_N) = E[X]$$

Consider a binomial tree with $p = 2/3$.

$$E_0[X_1] = \frac{4}{9}\cdot 1 + \frac{2}{9}\cdot 1 + \frac{2}{9}\cdot(-1) + \frac{1}{9}\cdot(-1) = \frac{1}{3}$$

$$E_0[X_2] = \frac{4}{9}\cdot 2 + 2\cdot\frac{2}{9}\cdot 0 + \frac{1}{9}\cdot(-2) = \frac{2}{3}$$

$$E_1[X_2] = \frac{2}{3}\cdot 2 + \frac{1}{3}\cdot 0 = \frac{4}{3}, \text{ if } X_1 = 1$$

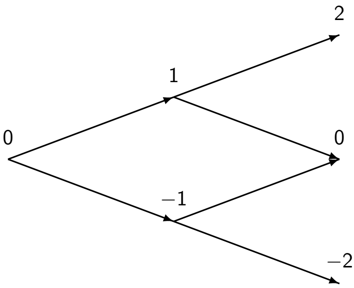
$$E_1[X_2] = \frac{2}{3}\cdot 0 + \frac{1}{3}\cdot(-2) = -\frac{2}{3}, \text{ if } X_1 = -1$$

Tower property:

$$E_0[E_1[X_2]] = \frac{2}{3}\cdot\frac{4}{3} + \frac{1}{3}\cdot\frac{-2}{3} = \frac{2}{3} = E_0[X_2]$$

Take out what is known:

$$E_1[X_1X_2] = X_1E_1[X_2] = \begin{cases} 1\cdot\frac{4}{3} = \frac{4}{3}, \text{ if } X_1 = 1 \\ -1\cdot\frac{-2}{3} = \frac{2}{3}, \text{ if } X_1 = -1 \end{cases}$$



In mathematical finance, we typically work within the framework of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

We normally have a fixed time T and a filtration, which is a collection of algebras $\mathcal{F}(t); 0 \leq t \leq T$ indexed by time variable t. We interpret $\mathcal{F}(t)$ as the information available at time t.

Conditioning on the information can be very helpful: in the binomial tree, for example, having knowledge on which events in $\mathcal{F}_1$ occurred and which not, gives full knowledge about the stock price at time 1.

Then, an **adapted stochastic process** (to be precise: adapted to filtration $\mathcal{F}(t)$) is a collection of random variables $\{X(t); 0 \le t \le T\}$ such that for every t, $X(t)$ is $\mathcal{F}(t)$-measurable.

This means that the information at time t is sufficient to evaluate the random variable $X(t)$.

We think of $X(t)$ as the price of some asset at time t and $\mathcal{F}(t)$ as the information obtained by watching all the prices in the market up to time t.

Definition

A stochastic process X_n is called an $(\mathcal{F}_n)$-**martingale** if

- X_n depends only on the first n coin tosses, i.e. is adapted to filtration $\mathcal{F}_n$.
- For all $m \leq n$

$$E_m[X_n] = X_m$$

"The best estimate for X_n is its current value X_m".

The process X_t is not necessarily the stock price S_t, but it depends on the same past coin tosses.

Consider a binomial tree with $p = 2/3$.

$$E_0[X_2] = \frac{2}{3} \quad E_0[X_1] = \frac{1}{3}$$

$$E_1[X_2] = \begin{cases} \frac{4}{3}, & \text{if } X_1 = 1 \\ -\frac{2}{3}, & \text{if } X_1 = -1 \end{cases}$$

$$E_0[X_2] \neq X_0, \quad E_0[X_1] \neq X_0$$

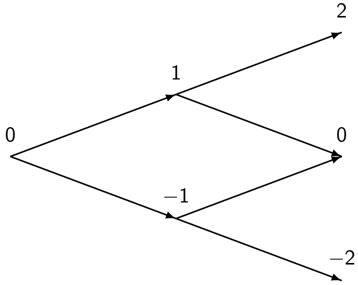
$$E_1[X_2] \neq X_1$$

Is X_t a martingale?

What if $p = 1/2$?

$$E_0[X_2] = E_0[X_1] = 0 = X_0 \quad \text{Check!}$$

$$E_1[X_2] = \begin{cases} 1 = X_1, & \text{if } X_1 = 1 \\ -1 = X_1, & \text{if } X_1 = -1 \end{cases} \qquad \text{Check!}$$



- Let's look back at the pricing formula in our binomial tree. We saw that in every step we had the following relation

$$V_n = \mathbb{E}_n^Q\left[\frac{V_{n+1}}{1+r}\right].$$

- Let's re-write it as

$$\frac{V_n}{(1+r)^n} = \mathbb{E}_n^Q\left[\frac{V_{n+1}}{(1+r)^{n+1}}\right].$$

- This shows that the following process

$$X_t = \frac{V_t}{(1+r)^t}$$

is a *martingale* under the risk-neutral measure.

A stochastic process X_n adapted to filtration $\mathcal{F}_n$ (i.e. depends only on the first n coin tosses) is a **Markov process** if the distribution of X_{n+1} conditioned on $\mathcal{F}_n$ is the same as the distribution of X_{n+1} conditioned on X_n.

- In other words, the future behavior of X_n does not depend on how the process arrived to this point, but depends only on the current value.
- This is a reflection of the belief that "markets have very short memory".

In the three-period binomial model we saw earlier this lecture:

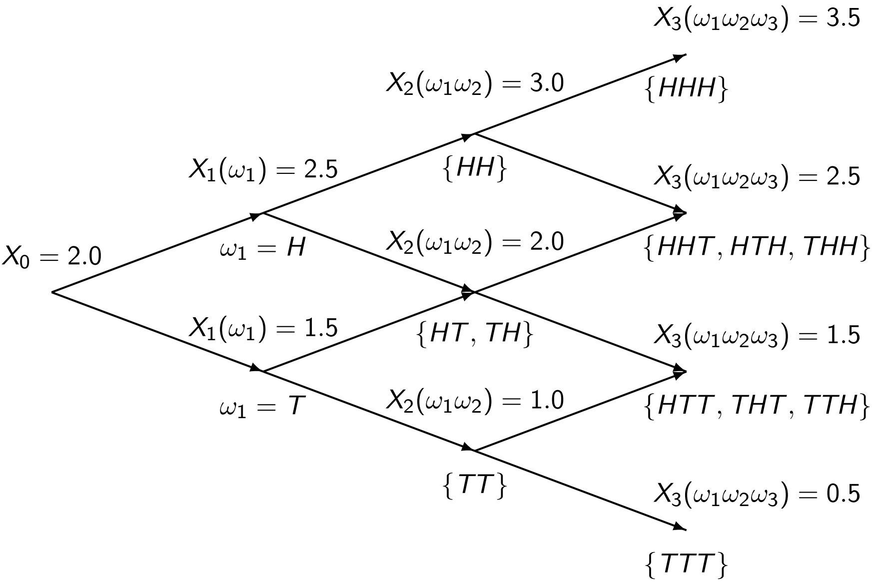
- We can arrive at $S_2^1 = sud$ in two different ways
- First heads and then tails or first tails and then heads
- However for the future behavior of S the order is irrelevant
- It only matters that the first two coin tosses generate one heads and one tails

Hence, in this model S is a Markov process.

Clip 2: From Finite to Infinite Sample Spaces

Mark-Jan Boes

September, 2020



$X_3(\omega_1\omega_2\omega_3) = 3.5$
$X_2(\omega_1\omega_2) = 3.0$
$\{HHH\}$
$X_1(\omega_1) = 2.5$
$\{HH\}$
$X_3(\omega_1\omega_2\omega_3) = 2.5$
$X_0 = 2.0$
$\omega_1 = H$
$X_2(\omega_1\omega_2) = 2.0$
$\{HHT, HTH, THH\}$
$X_1(\omega_1) = 1.5$
$\{HT, TH\}$
$X_3(\omega_1\omega_2\omega_3) = 1.5$
$\omega_1 = T$
$X_2(\omega_1\omega_2) = 1.0$
$\{HTT, THT, TTH\}$
$\{TT\}$
$X_3(\omega_1\omega_2\omega_3) = 0.5$
$\{TTT\}$

Recall the definition of an algebra: Let Ω be a nonempty set. A family $\mathcal{F}$ of subsets of Ω is an **algebra** if the following holds:

1. $\emptyset \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. $A_i \in \mathcal{F} \quad \text{for} \quad i = \overline{1, N} \quad \Rightarrow \quad \bigcup_i^N A_i \in \mathcal{F}$

In our tree model we have:

$$\Omega = \{\omega_1\omega_2\omega_3\} = \\ = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

It is easy to verify that an example of an algebra is $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

We can define probabilities for the elements in $\mathcal{F}_0$:

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(\Omega) = 1$

The whole idea is to define probabilities on all sets in an algebra.

Another example of an algebra in our 3-period tree model is $\mathcal{F}_1 = \{\emptyset, \Omega, A_1, A_2\}$, where:

$$A_1 = \{HHH, HHT, HTH, HTT\}$$
$$A_2 = \{THH, THT, TTH, TTT\}.$$

We now define the probabilities for the sets A_1 and A_2:

- $\mathbb{P}(A_1) = p$
- $\mathbb{P}(A_2) = q = 1 - p$

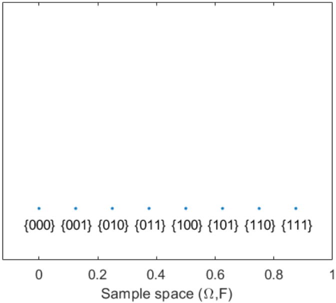
We have now defined $\mathbb{P}$ for the four sets in algebra $\mathcal{F}_1$.

Let us draw analogy with binary numbers and put $H = 1$ and $T = 0$.

Then, we can uniquely link the paths to numbers in the following way:

$$\{\omega_1\omega_2\omega_3\} \rightarrow \omega_1 \cdot 2^{-1} + \omega_2 \cdot 2^{-2} + \omega_3 \cdot 2^{-3}$$

For example: $\{000\} \rightarrow 0.0$, or $\{100\} \rightarrow 0.5$, or $\{011\} \rightarrow 0.375$.



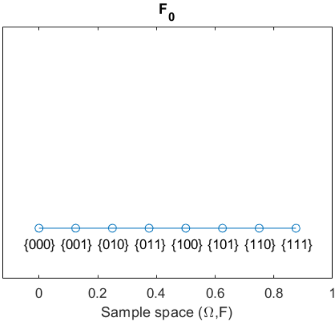
The graph visualizes our sample space Ω and filtration of algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3$

Let us draw analogy with binary numbers and put $H = 1$ and $T = 0$.

Then, we can uniquely link the paths to numbers in the following way:

$$\{\omega_1\omega_2\omega_3\} \to \omega_1 \cdot 2^{-1} + \omega_2 \cdot 2^{-2} + \omega_3 \cdot 2^{-3}$$

For example: $\{000\} \to 0.0$, or $\{100\} \to 0.5$, or $\{011\} \to 0.375$.



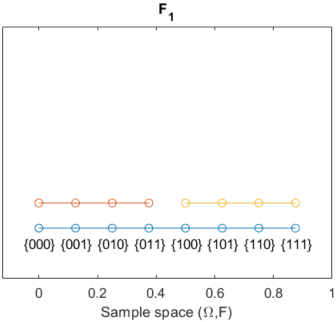
The graph visualizes our sample space Ω and filtration of algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3$

Let us draw analogy with binary numbers and put $H = 1$ and $T = 0$.

Then, we can uniquely link the paths to numbers in the following way:

$$\{\omega_1\omega_2\omega_3\} \to \omega_1 \cdot 2^{-1} + \omega_2 \cdot 2^{-2} + \omega_3 \cdot 2^{-3}$$

For example: $\{000\} \to 0.0$, or $\{100\} \to 0.5$, or $\{011\} \to 0.375$.



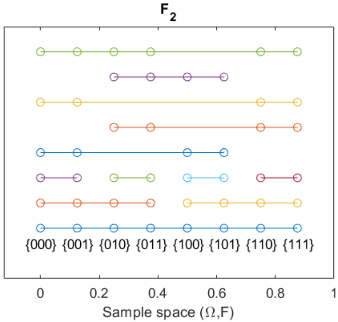
The graph visualizes our sample space Ω and filtration of algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3$

Let us draw analogy with binary numbers and put $H = 1$ and $T = 0$.

Then, we can uniquely link the paths to numbers in the following way:

$$\{\omega_1\omega_2\omega_3\} \to \omega_1 \cdot 2^{-1} + \omega_2 \cdot 2^{-2} + \omega_3 \cdot 2^{-3}$$

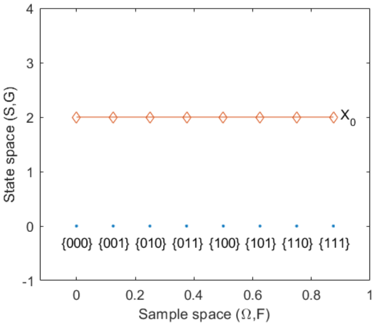
For example: $\{000\} \to 0.0$, or $\{100\} \to 0.5$, or $\{011\} \to 0.375$.



The graph visualizes our sample space Ω and filtration of algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3$

Next step is to define **random variables** on the measurable space $(\Omega, \mathcal{F})$.

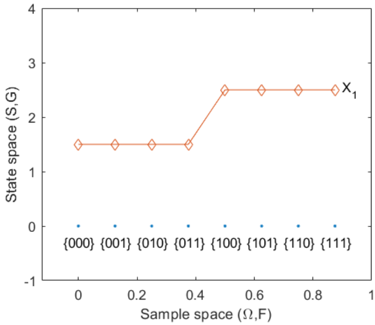
Remember how the spot price X_i depends on a particular path in the tree picture that we saw at the start of this lecture.



State space S of X_0 consists of a single value.

Next step is to define **random variables** on the measurable space $(\Omega, \mathcal{F})$.

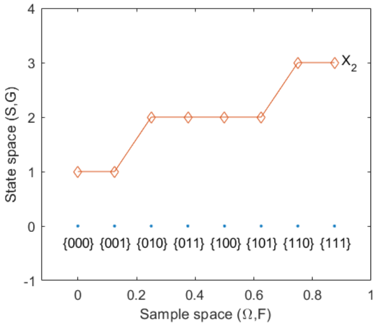
Remember how the spot price X_i depends on a particular path in the tree picture that we saw at the start of this lecture.



State space S of X_1 consists of two values.

Next step is to define **random variables** on the measurable space $(\Omega, \mathcal{F})$.

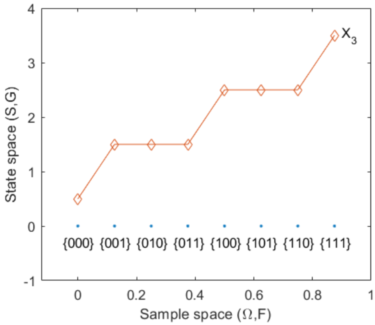
Remember how the spot price X_i depends on a particular path in the tree picture that we saw at the start of this lecture.



State space S of X_2 consists already of three values.

Next step is to define **random variables** on the measurable space $(\Omega, \mathcal{F})$.

Remember how the spot price X_i depends on a particular path in the tree picture that we saw at the start of this lecture.

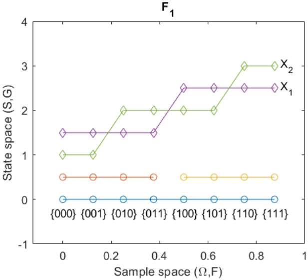


State space S of X_3 consists of four values.

So, we have looked at what value a random variable has for a particular element $\{\omega_1\omega_2\omega_3\}$ in sample space Ω.

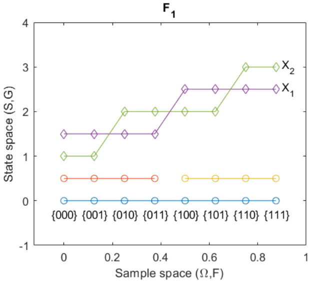
Let us look in the opposite direction: if we take a particular value or a subset in state space S, what does it correspond to in sample space Ω?

- This is called a **pre-image**.



- One can see that the pre-image of $X_1 = 1.5$ are the first four paths, the pre-image of $X_1 = 2.5$ are the last four paths.
- Both subsets are in $\mathcal{F}_1$.
- We say that X_1 is measurable with respect to $\mathcal{F}_1$

Again the picture:



What about X_2?

- The pre-image of $X_2 = 1$ are the first two paths
- This set is not in $\mathcal{F}_1$
- X_2 is **not** measurable with respect to $\mathcal{F}_1$

The intuition is pretty clear:

- Algebra $\mathcal{F}_1$ provides information on the outcome of the first coin toss
- Having knowledge on this outcome does not provide sufficient information on the value of X_2
- If tails comes up then X_2 could be either 1 or 2
- So, the information generated by $\mathcal{F}_1$ is insufficient to measure X_2

It is obvious that X_2 is measurable with respect to $\mathcal{F}_2$ (by construction):

- Note that the information revealed by X_2 is less than the information generated by $\mathcal{F}_2$
- Knowing the value of X_2 does not permit us to distinguish between an initial head followed by a tail from an initial tail followed by a head
- There is enough information in $\mathcal{F}_2$ to determine the value of X_2 and even more.

In general, a random variable X is **measurable** with respect to algebra $\mathcal{G}$ ($\mathcal{G}$-measurable) if and only if the information in $\mathcal{G}$ is sufficient to determine the value of X.

Up to now, we only dealt with discrete random variables.

One way to move to a continuous framework is to make more and more steps on our tree:

- Indeed, our mapping $\{\omega_1, \omega_2, \dots \omega_N\} \to \omega_1 \cdot 2^{-1} + \omega_2 \cdot 2^{-2} + \cdots + \omega_N \cdot 2^{-N}$ will be more and more dense on the interval $[0, 1]$.
- However, there is a crucial difference between this extended space Ω_N of random steps and e.g. the interval of real numbers $[0, 1]$: Ω_N is **countable**, while an interval of real numbers contains an **uncountable** number of points

Uncountable versus countable infinity:

- A set is **countably infinite** if its elements can be put in one-to-one correspondence with the set of natural numbers
- The set of integers $\mathcal{Z} = \{0, 1, -1, 2, -2, 3, -3, ...\}$ is clearly infinite
- Counting off the numbers in this set will take forever
- However, if you specify any integer, e.g. 10,458,274,805 you will get to this integer in a finite amount of time
- This makes it a **countably infinite** set
- A set that is **uncountably infinite** contains more elements than the infinite number of natural numbers
- For example the set of all real number between 0 and 0.001 is already uncountably infinite

At the start of his book Shreve makes the jump to uncountably infinite spaces by considering two experiments:

- choose a number from the unit interval [0,1], and
- toss a coin infinitely many times

For the second experiment, he defines

$$\Omega_\infty = \text{the set of infinite sequences of } H\text{s and } T\text{s.}$$

A generic element of Ω_∞ will be denoted $\omega = \omega_1\omega_2, \ldots,$ where ω_n indicates the result of the nth coin toss.

Issues with an uncountably infinite sample space:

- The probability of any particular outcome is zero
- Consequently, we are not able to determine the probability of a subset A of the sample space by summing the probabilities of all elements in A (like we did in the setting of a finite sample space!)
- Alternatively, we should define the **probability of subsets**
- However, in infinite sample spaces there are infinitely many events
- Solution: define the probability of some simple subsets and apply the properties of probabilities to determine the probability of more complicated subsets

In the next clip this will be illustrated for the random experiment where we toss a coin infinitely many times.

Clip 3: Probability and Measure in Infinite Sample Spaces

Mark-Jan Boes

September, 2020

A **finite probability space** consists of three elements:

- Ω, a nonempty set, called the sample space which contains all possible outcomes of a random experiment
- $\mathcal{F}$, an algebra of subsets of Ω, the subsets are called events
- $\mathbb{P}$, a probability measure on $(\Omega, \mathcal{F})$, i.e. $\mathbb{P}$ is function that assigns to each element $\omega \in \Omega$ a number in $[0, 1]$ such that

$$\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1.$$

An event is a subset of Ω and the probability of an event A is

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega).$$

Note that it follows from our definition that if subset A and subset B do not have common elements, i.e. $A \cap B = \emptyset$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

At the start of his book Shreve makes the jump to uncountably infinite spaces by considering two experiments:

- choose a number from the unit interval [0,1], and
- toss a coin infinitely many times

For the second experiment, he defines

$$\Omega_\infty = \text{the set of infinite sequences of } H\text{s and } T\text{s.}$$

A generic element of Ω_∞ will be denoted $\omega = \omega_1\omega_2, \ldots,$ where ω_n indicates the result of the nth coin toss.

Issues with an uncountably infinite sample space:

- The probability of any particular outcome is zero
- Consequently, we are not able to determine the probability of a subset A of the sample space by summing the probabilities of all elements in A (like we did in the setting of a finite sample space!)
- Alternatively, we should define the **probability of subsets**
- However, in infinite sample spaces there are infinitely many events
- Solution: define the probability of some simple subsets and apply the properties of probabilities to determine the probability of more complicated subsets

We generalize the definition of an **algebra** to a setting of an **infinite sample space**.

A family $\mathcal{F}$ of subsets of Ω is a σ-**algebra** if the following hold:

1. $\emptyset \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. $A_i \in \mathcal{F} \quad \text{for} \quad i = 1, 2, \ldots \quad \Rightarrow \quad \bigcup_i^\infty A_i \in \mathcal{F}$

Let Ω be a nonempty set and let $\mathcal{F}$ be a σ-algebra of subsets of Ω. A probability measure $\mathbb{P}$ is a function that, to every set $A \in \mathcal{F}$, assigns a number in $[0,1]$, called the probability of A and written $\mathbb{P}(A)$. We require:

- $\mathbb{P}(\Omega) = 1$, and
- whenever, $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{F}$, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.

Suppose we toss a coin infinitely many times.

Define:

$$\Omega_\infty = \text{the set of infinite sequences of } H\text{s and } T\text{s.}$$

Other assumptions:

- the probability of a head on each toss is $p > 0$
- the probability of a tail on each toss is $q = 1 - p > 0$
- coin tosses are independent

Purpose: we want to construct a probability measure for this random experiment.

Let us first consider σ-algebra:

$$\mathcal{F}_0 = \{\emptyset, \Omega\}.$$

We define:

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(\Omega) = 1$

We next define $\mathbb{P}$ for the next two sets:

$$A_H = \text{set of all sequences starting with} H$$
$$A_T = \text{set of all sequences starting with} T$$

by setting:

- $\mathbb{P}(A_H) = p$
- $\mathbb{P}(A_T) = q$.

We have now defined $\mathbb{P}$ for four sets.

We next define $\mathbb{P}$ for the next four sets:

$$\begin{aligned}
A_{HH} &= \text{the set of all sequences beginning with } HH \\
A_{HT} &= \text{the set of all sequences beginning with } HT \\
A_{TH} &= \text{the set of all sequences beginning with } TH \\
A_{TT} &= \text{the set of all sequences beginning with } TT,
\end{aligned}$$

by setting:

$$\mathbb{P}(A_{HH}) = p^2, \quad \mathbb{P}(A_{HT}) = pq, \quad \mathbb{P}(A_{TH}) = pq, \quad \mathbb{P}(A_{TT}) = q^2.$$

With this definition, we are able to determine the probability of all sets in $\mathcal{F}_2$, the σ-algebra that consists of the sets above and their complements and unions.

Continuing this process leads to the definition of every set that can be described in terms of finitely many coin tosses.

Once the probabilities of all these sets are specified, there are other sets, not describable in terms of finitely many coin tosses, whose probabilities should be determined.

For example, the set containing only the single sequence $HHHH\ldots$ cannot be described in terms of finitely many coin tosses, but it is a subset of A_H, A_{HH}, A_{HHH} etc.

Furthermore,

$$\mathbb{P}(A_H) = p, \quad \mathbb{P}(A_{HH}) = p^2, \quad \mathbb{P}(A_{HHH}) = p^3, \ldots,$$

Since these probability converge to zero, we **must** have

$$\mathbb{P}(\text{Every toss results in heads}) = 0.$$

Using the same argument we can show that every individual sequence in Ω_∞ has probability zero.

We create a σ-algebra $\mathcal{F}_\infty$ by putting in every set that can be described in terms of finitely many coin tosses and then adding all other sets required to have a σ-algebra.

It turns out that once the probability of every set that can be described in terms of finitely many coin tosses is specified, the probability of every set in $\mathcal{F}_\infty$ is determined.

The example shows that **every individual sequence** has probability zero.

It would be satisfying if we could say that events that have probability zero cannot happen.

For example: we would like to say that if we toss a coin infinitely many times, it cannot happen that we get a head on each toss. At the same time: we are **sure** to get at least one tail.

The peculiarity of our uncountably infinite sample space is that the sequence that is all heads is no less likely to happen than any other particular sequence.

So, mathematicians have created the terminology 'almost surely': whenever a event is said to be almost sure, we mean it has probability one, even though it may not include every possible outcome in the sample space.

Suppose we construct a mathematical model for choosing a number at random from the unit interval $[0, 1]$ such that the probability is distributed uniformly over this interval. We define the probability of closed intervals $[a, b]$ by the formula

$$\mathbb{P}[a, b] = b - a, \quad 0 \leq a \leq b \leq 1.$$

This particular probability measure on $[0, 1]$ is called Lebesgue measure: the Lebesgue measure of a subset in $\mathbb{R}$ is its 'length'.

Implication: if $b = a$ then $[a, b]$ is the set only containing the number a, the probability of this set is zero.

Another implication:

$$\mathbb{P}(a, b) = b - a.$$

There are many other subsets that are defined by this definition (and the definition of probability measure), for example the set:

$$\left[0, \frac{1}{3}\right] \bigcup \left[\frac{2}{3}, 1\right],$$

is not a closed interval but we can derive that its probability is $\frac{2}{3}$.

It is natural to ask whether it is possible to describe the collection of **all sets** whose probability is determined by the Lebesgue probability measure.

It turns out that this collection of sets is the σ-algebra we get starting with the closed intervals and putting in everything else required to have a σ-algebra: the Borel σ-algebra of subsets of $[0, 1]$.

These are the subsets of $[0, 1]$, the so-called events, whose probability is determined once we specify the probability of the closed intervals.

Let us look on some properties of the Borel algebra:

- Borel algebra includes single points. Indeed, define interval

$$X_n = [a - 1/n, a + 1/n].$$

  Intersection

$$a = \bigcap_{i=1}^{\infty} X_n$$

  converges to a single point a.

- As points belong to the Borel algebra, so do open intervals

$$\bigcup_{i=1}^{\infty} [a + 1/n, b - 1/n] \to (a, b).$$

Now, we are able to give a more general definition of a random variable:

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A **random variable** is a real-valued function X defined on Ω with the property that for every Borel subset B of $\mathbb{R}$, the subset of Ω given by

$$\{X \in B\} = \{\omega \in \Omega; X(\omega) \in B\},$$

is in the σ-algebra $\mathcal{F}$.

In words: a random variable X assigns a numerical value to each outcome of a random experiment or the random variable X is $\mathcal{F}$-measurable.

We are interested in determining $\mathbb{P}\{X \in B\}$, which is defined, given that the requirement that $\{X \in B\}$ is part of $\mathcal{F}$.

If we look at the tree that we used in the previous knowledge clip:

$$\mathbb{P}\{X_2 = 2\} = \mathbb{P}\{\omega \in \Omega; X_2(\omega) = 2\} = \mathbb{P}\{A_{HT} \cup A_{TH}\} = 2pq.$$

From a conceptual point of view it is important to keep in mind that a probability measure is defined on $\mathcal{F}$.

In words:

- We determine the events that lead to a value of 2 for random variable X_2: in this case it is the single event that the first two coin tosses have different outcomes
- By definition of the random variable we know that these events are in $\mathcal{F}$
- The probabilities of the events in $\mathcal{F}$ are measured with the probability measure $\mathbb{P}$

A finite **measure** μ on a measurable space $(Y, \mathcal{F})$, with Y a set and $\mathcal{F}$ an algebra, is a mapping

$$\mu : \mathcal{F} \to \mathbb{R}_+,$$

such that

1. $$\mu(A) \geq 0, \forall A \in \mathcal{F}.$$
2. $$\mu(\emptyset) = 0.$$
3. If $A_n \in \mathcal{F} \quad \forall n = 1, 2, \ldots$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

The triplet $(Y, \mathcal{F}, \mu_X)$ is called a **measure space**.

Hence, there is a difference between a measure space and a probability space.

A **probability space** is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$, where the measure $\mathbb{P}$ has the property:

$$\mathbb{P}(\Omega) = 1.$$

As we know by now: The space Ω is called a **sample space** and the subsets of the **sigma-algebra** $\mathcal{F}$ are called **events**.

And to be fully complete:

A **random variable** X on a measurable space is $(\Omega, \mathcal{F})$ a mapping

$$X : \Omega \to \mathbb{R}$$

such that X is $\mathcal{F}$-measurable.

Here, $\mathbb{R}$ is the **state space** of X.

We regularly use in calculations the probability density functions and cumulative distribution functions. These functions are defined on the **state space**, rather than on the **sample space**.

The **distribution measure** μ_X for a random variable X is the probability measure that assigns to each Borel subset B of $\mathbb{R}$ the mass.

$$\mu_X(B) = \mathbb{P}\{\omega \in \Omega; X(\omega) \in B\}, \quad B \in \mathcal{B},$$

i.e.

$$\mu_X(B) = \mathbb{P}(X^{-1}(B)).$$

Recall that $\mathcal{B}$ is the σ-algebra that is generated by the closed intervals in $\mathbb{R}$.

The cumulative **distribution function** of X is denoted by F_X and defined by

$$F_X(x) = \mathbb{P}\{\omega \in \Omega; X(\omega) \leq x\}.$$

So, probability measure is a probability of an arbitrary event, while the distribution function is a probability of the random variable to be below a specific level.

A stochastic process X on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a mapping

$$X : R^+ \times \Omega \to R,$$

such that for each $t \in R^+$ the mapping

$$X(t, \cdot) : \Omega \to R,$$

is $\mathcal{F}$-measurable.

Here, $R^+ \times \Omega$ is a product space, containing all combinations of elements in R^+ and Ω.

$X(t, \omega)$ is the value at time t given the outcome ω.

For each t, the mapping

$$\omega \to X(t, \omega),$$

is a random variable.

For each $\omega \in \Omega$ the mapping

$$t \to X(t, \omega)$$

is a deterministic function of time, often called a realization or trajectory of X.

We are typically interested in some parameters of random variables, mainly expected value and variance. We need proper definitions for these parameters.

For a variable X defined on $(\Omega, \mathcal{F}, \mathbb{P})$, the expectation of X (where Ω is finite) is defined as:

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega).$$

We typically compute the expectation of a random variable on the state space.

Next clip we will define expectation in case Ω is uncountable infinite.

I hear you thinking: what is it what I actually learned from this clip?

Fair question: the stuff that we have been going through is rather abstract and formal.

And, frankly: you have a pretty good notion of what a process is in general, you don't need the material of the past two clips for that.

You can pretty well imagine that the path a stock price is a stochastic process: state space, sample spaces and algebras are not necessary to know that.

So, why do we spend so much time on this?

- As quantitative oriented students it is good to have a basic knowledge on how mathematicians formally think of probability and measure
- Last week we have seen that one way to determine the no-arbitrage price of a derivative is to maintain the sample space but change the probabilities: formally this is a change of probability measure
- One lesson of this clip is that from a mathematical point of view a change of (probability) measure is nothing fancy: you can define a lot of different (probability) measures on the same measure space

Clip 4: Integration and Expectation

Mark-Jan Boes

September, 2020

We are typically interested in some parameters of random variables, mainly expected value and variance. We need proper definitions for these parameters.

For a random variable X defined on a **finite probability space** $(\Omega, \mathcal{F}, \mathbb{P})$, the expectation of X is defined as:

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega).$$

If Ω is countably infinite, its elements can be listed in a sequence $\omega_1\omega_2\omega_3, \ldots,$ and we can define $\mathbb{E}(X)$ as an infinite sum:

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} X(\omega_k)\mathbb{P}(\omega_k).$$

Difficulties arise if Ω is **uncountably** infinite: uncountable sums cannot be defined. Instead, we must think in terms of **integrals**.

You remember how the **Riemann integral** is defined? No, let me help you out.

If $f(x)$ is a continuous function defined for all x in the closed interval $[a, b]$, the **Riemann integral** $\int_a^b f(x)dx$ is defined as follows.

First partition $[a, b]$ into subintervals

$$[x_0, x_1], [x_1, x_2], ..., [x_{n-1}, x_n],$$

where $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$.

We denote by $\Pi = \{x_0, x_1, ..., x_n\}$ the set of partition points and by

$$||\Pi|| = \max_{1 \leq k \leq n} (x_k - x_{k-1}),$$

the length of the longest subinterval in the partition. For each subinterval $[x_{k-1}, x_k]$, we set

$$M_k = \max_{x_{k-1} \leq x \leq x_k} f(x),$$

and

$$m_k = \min_{x_{k-1} \leq x \leq x_k} f(x).$$

The upper Riemann sum is:

$$\mathrm{RS}_{\Pi}^{+} = \sum_{k=1}^{n} M_k (x_k - x_{k-1}),$$

and the lower Riemann sum:

$$\mathrm{RS}_{\Pi}^{-} = \sum_{k=1}^{n} m_k (x_k - x_{k-1}).$$

Graphically:

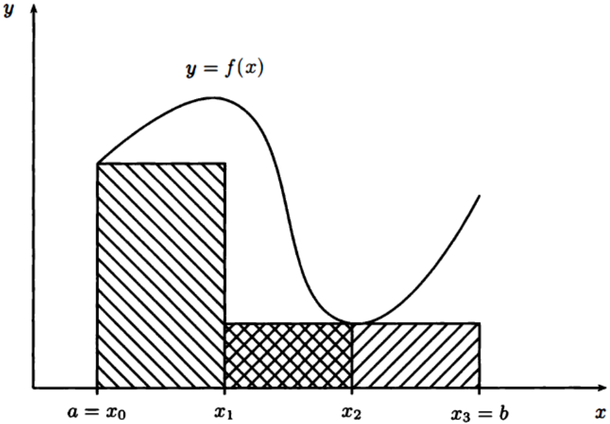


Fig. 1.3.1. Lower Riemann sum.

As $||\Pi||$ converges to zero, i.e. we put in more and more partition points and the subinterval in the partition become shorter and shorter, the upper Riemann sum and the lower Riemann sum converge to the same limit.

This limit is the so-called Riemann integral, which is denoted by:

$$\int_a^b f(x)dx.$$

Note that an unbounded function is not Riemann-integrable.

Why can't we imitate this procedure to define the expectation of a continuous random variable X?

- X is a function of $\omega \in \Omega$
- Ω is often not a subset in $\mathbb{R}$
- There is no natural way to partition the set Ω as we partitioned the interval $[a, b]$

The solution is to partition the y-axis.

Assume for now that $0 \leq X(\omega) < \infty$ for every $\omega \in \Omega$, and let $\Pi = \{y_0, y_1, y_2, ...\}$ where $0 = y_0 < y_1 < y_2 < ...$. For each subinterval $[y_k, y_{k+1}]$ we set

$$A_k = \{\omega \in \Omega; y_k \leq X(\omega) < y_{k+1}\}.$$

We define the lower Lebesgue sum to be:

$$LS_\Pi^- = \sum_{k=1}^{\infty} y_k \mathbb{P}(A_k).$$

Graphically:

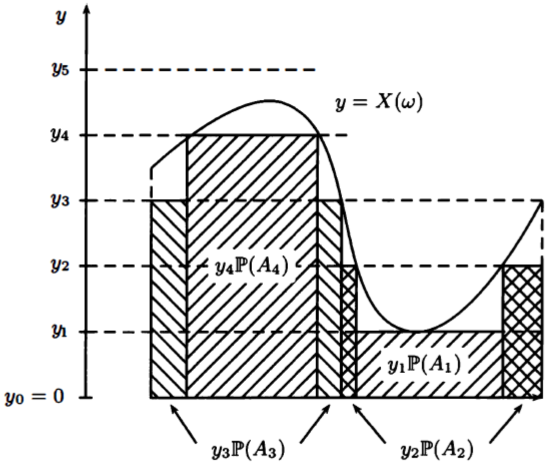


Fig. 1.3.2. Lower Lebesgue sum.

This lower sum converges as $||\Pi||$, the maximum distance between the y_k partition points, approaches zero.

We define this limit to be the Lebesgue integral:

$$\int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

The Lebesgue integral might be ∞ because we have not made any assumptions on how large the values of X can be.

We assumed $0 \leq X(\omega) < \infty$ for every $\omega \in \Omega$. If the set of ω that violates this condition has zero probability, there is no effect on the integral. Otherwise, we define:

$$\int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \infty.$$

Finally, we need to look at random variables X that can take both positive and negative values. We define the positive and negative parts of X:

$$X^{+}(\omega) = \max(X(\omega), 0) \quad X^{-}(\omega) = \max(-X(\omega), 0).$$

Note that both X^{+} and X^{-} are nonnegative random variables, and

$$X = X^{+} - X^{-}.$$

Provided that both parts are not ∞, we can define:

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) = \int_\Omega X^+(\omega)d\mathbb{P}(\omega) - \int_\Omega X^-(\omega)d\mathbb{P}(\omega).$$

If both parts on the RHS are finite we say that X is integrable, and $\int_\Omega X(\omega)d\mathbb{P}(\omega)$ also finite.

In case both parts on the RHS are infinite then $\int_\Omega X(\omega)d\mathbb{P}(\omega)$ is not defined.

Formal definition of expectation:

Let X be a random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The expectation (or expected value) of X is defined to be:

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

The idea is to average the values of $X(\omega)$ over Ω, taking the probabilities into account.

This definition makes sense if X is integrable. We say that a random variable X is integrable if and only if:

$$\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) < \infty.$$

We often want to integrate a random variable X over a subset A of Ω rather than over all Ω. For this reason we define:

$$\int_A X(\omega) d\mathbb{P}(\omega) = \int_\Omega \mathbb{I}_A(\omega) X(\omega) d\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{F},$$

where $\mathbb{I}_A$ is the indicator function (random variable) given by:

$$\mathbb{I}_A(\omega) = \begin{cases} 1, & \text{if} \quad \omega \in A, \\ 0, & \text{if} \quad \omega \notin A. \end{cases}$$

If A en B are disjoint sets in $\mathcal{F}$ then

$$\int_{A \cup B} X(\omega) d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega) + \int_B X(\omega) d\mathbb{P}(\omega).$$

Still, the abstract space Ω is not a pleasant environment in which to actually compute integrals.

Recall that the distribution measure of X is the probability measure on $\mathbb{R}$ by

$$\mu_X(B) = \mathbb{P}\{\omega \in \Omega : X(\omega) \in B\}.$$

Because μ_X is a **probability measure** on $\mathbb{R}$, we can use it to integrate over $\mathbb{R}$:

$$\mathbb{E}|g(x)| = \int_{\mathbb{R}} |g(x)| d\mu_X(x).$$

This tells us that in order to compute the Lebesgue integral $\mathbb{E}(X) = \int_\Omega X(\omega) d\mathbb{P}(\omega)$ over the abstract space Ω, it suffices to compute $\int_\mathbb{R} g(x) d\mu_X(x)$.

In the simplest case X takes only finitely many values $x_0, x_1, x_2, \dots, x_n$ and then μ_X places a mass of $p_k = \mathbb{P}\{\omega \in \Omega : X(\omega) = x_k\}$ at each number x_k. Then:

$$\mathbb{E}(g(X)) = \int_\mathbb{R} g(x) d\mu_X = \sum_{k=0}^{n} g(x_k) p_k.$$

The most common case for continuous time models in finance is **when X has a density**. This means that there is a nonnegative, Borel-measurable function f defined on $\mathbb{R}$ such that

$$\mu_X(B) = \int_B f(x)dx \quad \text{for every Borel subset } B \text{ in } \mathbb{R}.$$

This density allows us to compute the measure μ_X of a set B by computing an integral over B.

This is the way you have learned to compute expectation in basic statistics courses:

$$\mathbb{E}|g(x)| = \int_{-\infty}^{\infty} |g(x)| f(x)dx.$$

Clip 5: Conditional Probability and Conditional Expectation

Mark-Jan Boes

September, 2020

The probability $\mathbb{P}(A)$ of an event A can be interpreted as a measure of the likelihood that A occurs.

If we have some additional information such as that another event has occurred, then our estimate of this likelihood may change.

For example in the 3-period binomial tree model:

- The probability of the event "three times tails" has non-zero probability
- However conditional on the event "heads at the first coin toss" this probability is zero

In general, **conditional probability** $\mathbb{P}(A|B)$ for the event A given that the event B has occurred is defined by the formula:

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

provided that $\mathbb{P}(B) > 0$.

Conditioning on event B means that we know that event B has happened. Hence, the space of of possible events is B rather than Ω. So, the weight of all possible event is now $\mathbb{P}(B)$ and not $\mathbb{P}(\Omega)$

The likelihood for the occurrence of an event could be **unaffected** by whether or not another event B has occurred. In such a case conditioning on B does not have an impact on the probability of A happening, i.e $\mathbb{P}(A|B) = \mathbb{P}(A)$.

This implies:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

We say that events A and B are **independent** if and only if this relation holds.

We want to introduce a definition of **conditional expectation** for stochastic processes

- Our **unconditional expectation** of a random variable X was defined as

$$\mathbb{E}(X) = \int_{\Omega} X(\omega)\mathbb{P}(d\omega).$$

- Now suppose we know that in some experiment the outcome ω is in event B.
- We can ask: what is the expectation of X given event B happened?
- Intuitively, we need to take an average over the effective sample space B. And we need to normalize by the weight of B.

Definition

Suppose $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$ and X is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the conditional expectation of X given B is defined as

$$\mathbb{E}[X|B] = \frac{1}{\mathbb{P}(B)} \int_B X(\omega) d\mathbb{P}(\omega).$$

Hence, we use the conditional probabilities to weigh the remaining possible outcomes of X.

Problem: in continuous time models the probability of an event is typically zero.

Let's try to get intuition for what we can do in the continuous case by looking at the discrete case.

Consider the general 3-period binomial model. The probability of head is p and of tail is $q = 1 - p$. Then

$$\begin{aligned}
\mathbb{E}_2(S_3)(HH) &= pS_3(HHH) + qS_3(HHT) \\
\mathbb{E}_2(S_3)(HT) &= pS_3(HTH) + qS_3(HTT) \\
\mathbb{E}_2(S_3)(TH) &= pS_3(THH) + qS_3(THT) \\
\mathbb{E}_2(S_3)(TT) &= pS_3(TTH) + qS_3(TTT).
\end{aligned}$$

On the previous slide we were conditioning on the information on the first two coin tosses: what is our best estimate of the stock price at time 3 *given* the outcomes of the first two coin tosses.

We multiply LHS with the probability of the event (e.g. in the first case the event that two heads come up):

$$\begin{aligned} p^2\mathbb{E}_2(S_3)(HH) &= p^2(pS_3(HHH) + qS_3(HHT)) \\ pq\mathbb{E}_2(S_3)(HT) &= pq(pS_3(HTH) + qS_3(HTT)) \\ pq\mathbb{E}_2(S_3)(TH) &= pq(pS_3(THH) + qS_3(THT)) \\ q^2\mathbb{E}_2(S_3)(TT) &= q^2(pS_3(TTH) + qS_3(TTT)). \end{aligned}$$

We can rewrite the RHS as follows:

$$p^2\mathbb{E}_2(S_3)(HH) = \sum_{\omega \in A_{HH}} S_3(\omega)\mathbb{P}(\omega)$$

$$pq\mathbb{E}_2(S_3)(HT) = \sum_{\omega \in A_{HT}} S_3(\omega)\mathbb{P}(\omega)$$

$$pq\mathbb{E}_2(S_3)(TH) = \sum_{\omega \in A_{TH}} S_3(\omega)\mathbb{P}(\omega)$$

$$q^2\mathbb{E}_2(S_3)(TT) = \sum_{\omega \in A_{TT}} S_3(\omega)\mathbb{P}(\omega).$$

In a discrete framework we could rewrite the LHS as follows:

$$\mathbb{E}_2(S_3)(HH)(\mathbb{P}(HHH) + \mathbb{P}(HHT)) = \sum_{\omega \in A_{HH}} S_3(\omega)\mathbb{P}(\omega)$$

$$\mathbb{E}_2(S_3)(HT)(\mathbb{P}(HTH) + \mathbb{P}(HTT)) = \sum_{\omega \in A_{HT}} S_3(\omega)\mathbb{P}(\omega)$$

$$\mathbb{E}_2(S_3)(TH)(\mathbb{P}(THH) + \mathbb{P}(THT)) = \sum_{\omega \in A_{TH}} S_3(\omega)\mathbb{P}(\omega)$$

$$\mathbb{E}_2(S_3)(TT)(\mathbb{P}(TTH) + \mathbb{P}(TTT)) = \sum_{\omega \in A_{TT}} S_3(\omega)\mathbb{P}(\omega).$$

On each of the sets where we condition on, the conditional expectation $\mathbb{E}_2(S_3)$ is constant because the conditional expectation does not depend on the third coin toss and the events are created by holding the first two coin tosses fixed.

Consequently we can write for continuous models:

$$\int_{A_{HH}} \mathbb{E}_2(S_3)(\omega) d\mathbb{P}(\omega) = \int_{A_{HH}} S_3(\omega) d\mathbb{P}(\omega)$$
$$\int_{A_{HT}} \mathbb{E}_2(S_3)(\omega) d\mathbb{P}(\omega) = \int_{A_{HT}} S_3(\omega) d\mathbb{P}(\omega)$$
$$\int_{A_{TH}} \mathbb{E}_2(S_3)(\omega) d\mathbb{P}(\omega) = \int_{A_{TH}} S_3(\omega) d\mathbb{P}(\omega)$$
$$\int_{A_{TT}} \mathbb{E}_2(S_3)(\omega) d\mathbb{P}(\omega) = \int_{A_{TT}} S_3(\omega) d\mathbb{P}(\omega)$$

.

We can generalize this by defining conditional expectation for every $A \in \mathcal{F}_2$ as:

$$\int_A \mathbb{E}_2(S_3)(\omega) d\mathbb{P}(\omega) = \int_A S_3(\omega) d\mathbb{P}(\omega).$$

In words: the value of the conditional expectation has been chosen to be that constant that yields the same average over the event as the random variable being estimated.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X a random variable. Let $\mathcal{G}$ be a sigma-algebra such that $\mathcal{G} \subseteq \mathcal{F}$. If Z is a random variable such that

1. Z is $\mathcal{G}$-measurable.
2. For every $B \in \mathcal{G}$ it holds that

$$\int_B Z(\omega) d\mathbb{P}(\omega) = \int_B X(\omega) d\mathbb{P}(\omega),$$

then we say the Z is the conditional expectation of X given the sigma-algebra $\mathcal{G}$ and we denote it as

$$\mathbb{E}[X|\mathcal{G}].$$

Clip 1: Random Walk

Mark-Jan Boes

September, 2020

Part I

Clip 1

Last week we have seen that in mathematical finance, we typically work within the framework of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

In this setup $\mathcal{F}$ is a collection of subsets of Ω, also called **events**.

A **stochastic process** X on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a mapping

$$X : \mathbb{R}^+ \times \Omega \to \mathbb{R},$$

such that for each $t \in \mathbb{R}^+$ the mapping

$$X(t, \cdot) : \Omega \to \mathbb{R},$$

is $\mathcal{F}$-measurable.

Here, $\mathbb{R}^+ \times \Omega$ is a product space, containing all combinations of elements in $\mathbb{R}^+$ and Ω.

So, in fact a stochastic process is a sequence of random variables through time.

But there is more to say. In our analysis, we normally have a fixed time T and a **filtration**, which is a **collection of σ-algebras** $\mathcal{F}(t); 0 \leq t \leq T$ indexed by time variable t. We interpret $\mathcal{F}(t)$ as the information available at time t.

Then, a **stochastic process adapted to filtration** $\mathcal{F}(t)$ is a collection of random variables $\{X(t); 0 \leq t \leq T\}$ such that for every t, $X(t)$ is $\mathcal{F}(t)$-measurable.

We think of $X(t)$ as the price of some asset at time t and $\mathcal{F}(t)$ as the information obtained by watching all the prices in the market up to time t.

We are interested in defining a process for the underlying value of a derivatives contract: with a properly defined process for the underlying value, we can determine the probabilistic behaviour of the payoff of the derivative and price the derivative payoff accordingly (see knowledge clips week 1).

The binomial tree model is very insightful but is too simple to serve as an adequate description of the real world dynamics of the underlying stock price: only two possible outcomes for the stock price.

However, it would already become much more realistic if we would split up the period in a number of smaller periods e.g. divide the year in 12 months.

Taking a time step of one day results in an option price that is very close to the famous Black-Scholes option pricing formula.

The **Black-Scholes model** is built around the following stochastic differential equation for the stock (index):

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

where W is a Brownian Motion under the real world probability measure $\mathbb{P}$.

It is a continuous time model: a **discrete time approximation** would look as follows:

$$S_{t+h} - S_t \approx \mu S_t h + \sigma S_t (W_{t+h} - W_t),$$

Question: how is the Brownian Motion process defined?

Brownian Motion is the **continuous time random walk**.

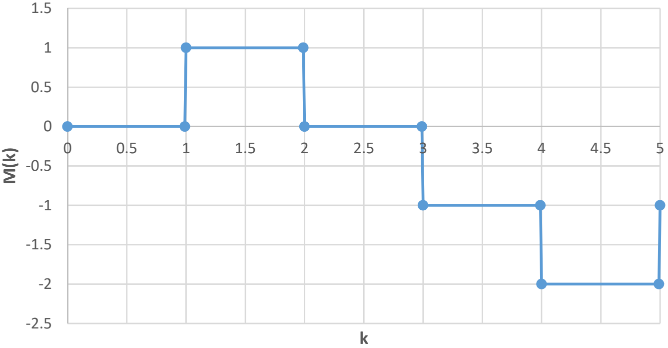
Consider:

$$X_j = \begin{cases} 1 & \text{if heads } (p = 0.5), \\ -1 & \text{if tails} \end{cases}$$

and define $M_0 = 0$, and

$$M_k = \sum_{j=1}^{k} X_j.$$

This is called a **symmetric random walk**:



Suppose we choose nonnegative integers $0 = k_0 < k_1 < \cdots < k_m$, the random variables

$$M_{k_1} = (M_{k_1} - M_{k_0}), (M_{k_2} - M_{k_1}), \cdots, (M_{k_m} - M_{k_{m-1}})$$

are independent. Define the increment as:

$$M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j.$$

This is the change in the position of the random walk between times k_i and k_{i+1}.

Increments over nonoverlapping time intervals are **independent** because they depend on different coin tosses.

Each increment has expectation equal to zero:

$$\mathbb{E}(M_{k_{i+1}} - M_{k_i}) = \mathbb{E}\left(\sum_{j=k_i+1}^{k_{i+1}} X_j\right) = \sum_{j=k_i+1}^{k_{i+1}} \mathbb{E}\left(X_j\right) = 0.$$

The variance of the increment is:

$$\begin{aligned}\mathrm{Var}(M_{k_{i+1}} - M_{k_i}) &= \mathrm{Var}\left(\sum_{j=k_i+1}^{k_{i+1}} X_j\right) = \sum_{j=k_i+1}^{k_{i+1}} \mathrm{Var}(X_j) = \sum_{j=k_i+1}^{k_{i+1}} 1 \\ &= k_{i+1} - k_i.\end{aligned}$$

The symmetric random walk is a martingale. To see this we choose nonnegative integers $k < l$ and compute:

$$
\begin{aligned}
\mathbb{E}\left[M_l|\mathcal{F}_k\right] &= \mathbb{E}\left[(M_l - M_k) + M_k|\mathcal{F}_k\right] \\
&= \mathbb{E}\left[(M_l - M_k)|\mathcal{F}_k\right] + \mathbb{E}\left[M_k|\mathcal{F}_k\right] \\
&= \mathbb{E}\left[(M_l - M_k)|\mathcal{F}_k\right] + M_k \\
&= \mathbb{E}\left[(M_l - M_k)\right] + M_k = M_k.
\end{aligned}
$$

Relating this to the previous clips:

- **Martingale**: the expectation of a future value of a random variable given today's information equals today's value of the random variable
- Hence, we calculate a **conditional expectation**
- We condition on σ-algebra $\mathcal{F}_k$, i.e. we have information on the first k coin tosses
- So, we have knowledge on $X_1, X_2, ..., X_k$ and therefore also on $M_1, M_2, ..., M_K$; X_k and M_k are $\mathcal{F}_k$-measurable
- The second equality is the result of the linearity of conditional expectations
- The third equality uses that M_k is $\mathcal{F}_k$-measurable, i.e. knowledge on the first k coin tosses defines the value of M_k
- The fourth equality follows from independence, $\mathcal{F}_k$ define M_k but does not say anything about the future

We consider the **quadratic variation** of the symmetric random walk, which is defined as (up to time k):

$$[M, M]_k = \sum_{j=1}^{k} (M_j - M_{j-1})^2 = \sum_{j=1}^{k} 1 = k.$$

Although the result is in this case the same as for $\mathrm{Var}(M_k)$, this is not true in general.

The key difference is that quadratic variation is calculated for a *path* while variance measures the centralized second moment of a random variable, i.e. probabilities are involved in the calculation.

Part II

Clip 2

In the previous knowledge clip we introduced the **symmetric random walk**.

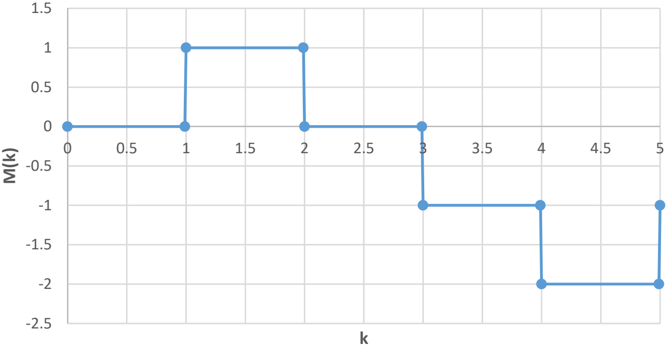
Consider:

$$X_j = \begin{cases} 1 & \text{if heads } (p = 0.5), \\ -1 & \text{if tails} \end{cases}$$

and define $M_0 = 0$, and

$$M_k = \sum_{j=1}^{k} X_j.$$

Graphical presentation of the **symmetric random walk**:



The increment was defined as:

$$M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j.$$

This is the change in the position of the random walk between times k_i and k_{i+1}.

Properties of increments:

- Increments over nonoverlapping time intervals are **independent**
- Increments have expectation equal to zero
- Increments have variance equal to the time span
- Increments have quadratic variation equal to the time span

To approximate a Brownian Motion process, we

- speed up time, i.e. we are going to toss the coin more often
- scale down the step size, i.e. the step size will be smaller than 1 (in absolute terms)

More precisely, we fix a positive integer n and define:

$$W^n(t) = \frac{1}{\sqrt{n}} M_{nt}$$

For example (from time $t = 1$ to $t = 4$):

$$W^{100}(4) = \frac{1}{\sqrt{100}} M_{400}$$

Hence, 400 coin tosses for one sample path with a step up or down of size $\frac{1}{10}$ on each coin toss.

One path of $W^{100}(4)$:

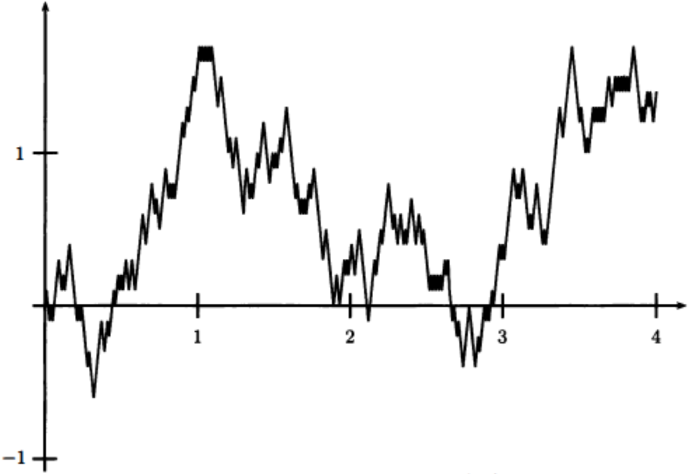


Fig. 3.2.2. A sample path of $W^{(100)}$.

Properties of the scaled symmetric random walk:

- independent increments
- the expectation of increments equals zero
- the variance of increments equals the time span
- the scaled symmetric random walk is a martingale
- the quadratic variation equals the time span

Hence, these properties **are the same as** for the symmetric random walk. However the random variable $W^{100}(4)$ has many more possible outcomes than M_4.

Let us look at the quadratic variation of the scaled random walk.

To be specific: the quadratic variation for W^{100} up to time 4.

$$
\begin{aligned}
\left[W^{100}, W^{100}\right](4) &= \sum_{j=1}^{400} \left[W^{100}\left(\frac{j}{100}\right) - W^{100}\left(\frac{j-1}{100}\right) \right]^2 \\
&= \sum_{j=1}^{400} \left[\frac{1}{10} X_j\right]^2 = \sum_{j=1}^{400} \frac{1}{100} = 4.
\end{aligned}
$$

We can fix t and think about the scaled random walk corresponding to different values of ω, the sequence of coin tosses.

E.g., set $t = 0.25$ and consider the possible set of values of:

$$W^{100}(0.25) = \frac{1}{\sqrt{100}} M_{25}.$$

This random variable is generated by 25 coin tosses and can take any of the values

$$-2.5, -2.3, -2.1, \cdots, -0.3, -0.1, 0.1, 0.3, \cdots, 2.1, 2.3, 2.5$$

So, we have a discrete sample space for $W^{100}(0.25)$.

In order for $W^{100}(0.25)$ to take the value of 0.1, we must get 13 heads and 12 tails in the 25 tosses.

What is the probability of this happening? To answer this question we use the pdf of the binomial distribution:

$$\mathbb{P}\left(W^{100}(0.25) = 0.1\right) = \frac{25!}{13!12!}\left(\frac{1}{2}\right)^{25} = 15.55\%.$$

In this way, the full probability distribution of $W^{100}(0.25)$ can be obtained.

What happens if $n \to \infty$?

Fix $t \geq 0$. As $n \to \infty$, the distribution of the scaled random walk $W^n(t)$ evaluated at time t converges to the normal distribution with mean zero and variance t.

So, what do we have in the limit:

- a **continuous time** stochastic process, a collection of random variables $\{W_t^\infty; 0 \leq t \leq T\}$.
- at each time point t, the random variable W_t^∞ has a normal distribution with expectation zero and variance t.

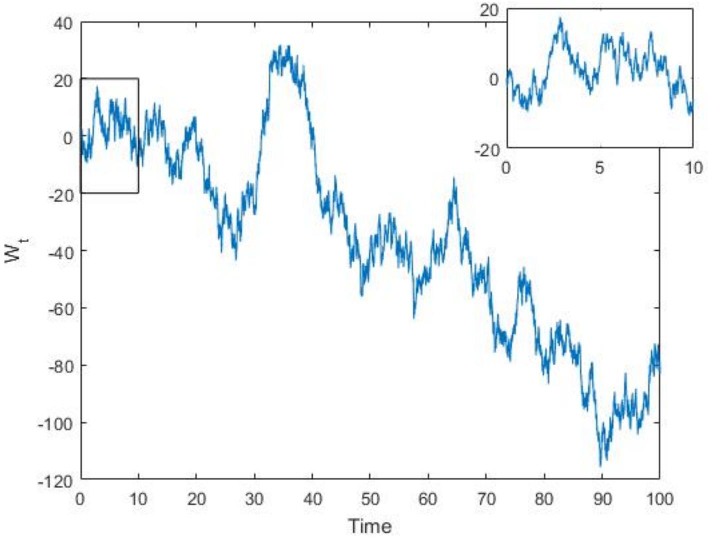
This is **Brownian Motion**: the limit of scaled random walks $W^n(t)$ as $n \to \infty$.

Formal definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose there is a continuous function $W(t)$ of $t \geq 0$ that satisfies $W(0) = 0$ and that depends on ω. Then $W(t)$, $t \geq 0$, is a **Brownian Motion** if for all $0 = t_0 < t_1 < \cdots < t_m$ the increments

$$W(t_1) = W(t_1) - W(t_0), W(t_2) - W(t_1), \cdots, W(t_m) - W(t_{m-1})$$

are **independent** and each of these increments is **normally distributed** with

$$\mathbb{E}\left[W_{t_{i+1}} - W_{t_i}\right] = 0$$
$$\mathrm{Var}\left[W_{t_{i+1}} - W_{t_i}\right] = t_{i+1} - t_i.$$


40
20
0
-20
-40
-60
-80
-100
-120
0
10
20
30
40
50
60
70
80
90
100
Time
W_t
20
0
-20
0
5
10

Part III

Clip 3

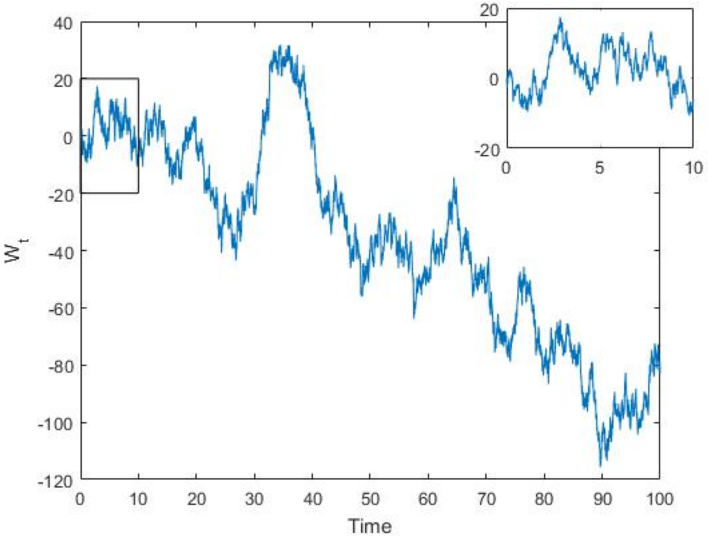
In the previous knowledge clip we introduced **Brownian Motion** as the limit of a scaled random walk $W^n(t)$ (with $n \to \infty$).

Formal definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose there is a **continuous function** $W(t)$ of $t \geq 0$ that satisfies $W(0) = 0$ and that depends on ω. Then $W(t)$, $t \geq 0$, is a **Brownian Motion** if for all $0 = t_0 < t_1 < \cdots < t_m$ the increments

$$W(t_1) = W(t_1) - W(t_0), W(t_2) - W(t_1), \cdots, W(t_m) - W(t_{m-1})$$

are **independent** and each of these increments is **normally distributed** with

$$\mathbb{E}\left[W_{t_{i+1}} - W_{t_i}\right] = 0$$
$$\mathrm{Var}\left[W_{t_{i+1}} - W_{t_i}\right] = t_{i+1} - t_i.$$



40
20
0
-20
-40
-60
-80
-100
-120
W_t
0
10
20
30
40
50
60
70
80
90
100
Time
20
0
-20
0
5
10

So, we had a scaled symmetric random walk: discrete time and a discrete number of possible outcomes at a particular time t.

The scaled symmetric random walk resulted from a finite number of coin tosses and a step size that depended on the number of coin tosses per unit of time.

Then we decided to toss the coin infinitely fast:

- leads to a continuous time stochastic process
- the sample space is also continuous: at a particular point in time, the process can take infinitely many values
- the probability distribution of the process at a particular point in time is the normal distribution

One of the key differences between Brownian Motion $W(t)$ and a scaled random walk, e.g. $W^{100}(t)$, is that the scaled random walk has a natural time step $\frac{1}{100}$ and is linear between these time steps whereas the Brownian Motion has no linear pieces.

Other difference: while the scaled random walk is only **approximately normal** (with sufficient steps for each unit of time) for each t, the Brownian Motion is **exactly normal**. This is the consequence of the central limit theorem.

Because the increments of Brownian Motion are independent and normally distributed, the random variables $W(t_1), W(t_2), \ldots, W(t_m)$ are jointly normally distributed.

For any two times $0 \le s < t$, the covariance of $W(s)$ and $W(t)$ is given by:

$$\begin{aligned} \mathrm{Cov}(W(s), W(t)) &= \mathbb{E}[(W(s)W(t)] - \mathbb{E}[W(s)]\mathbb{E}[W(t)] \\ &= \mathbb{E}[W(s)(W(t) - W(s)) + W^2(s)] \\ &= \mathbb{E}[W(s)]\mathbb{E}[W(t) - W(s)] + \mathbb{E}[W^2(s)] \\ &= 0 + \mathrm{Var}[W(s).] \end{aligned}$$

For the scaled symmetric random walk we could say:

$$\Delta W^n_{t_i} = W^n_{t_{i+1}} - W^n_{t_i}.$$

In the **continuous time framework**, we write:

$$dW(t).$$

Simply stated, Δ *becomes very small, infinitely small (infinitesimal)*.

Like Δ, d is a *forward looking operator*:

$$dW(t) = W(t+dt) - W(t), \quad \text{as is} \quad \Delta W^n_{t_i} = W^n(t_i+\Delta) - W^n(t_i)$$

So, $dW(t)$ is the increment of Brownian Motion over an infinitesimal timestep.

The increment $dW(t)$ is also a random variable:

- the distribution of $dW(t)$ is normal
- $\mathbb{E}_t[dW(t)] = 0$
- $Var_t[dW(t)] = dt$

We need to remember:

- $W(t) - W(s)$ has a Normal distribution with mean 0 and variance $t - s$, for any interval length $t - s$, $W(t) - W(s) \sim \mathcal{N}(0, t - s)$.
- What can we say about: $\int_s^t dW(u)$
- If $W(t)$ would be a differentiable function with respect to time then we could say

$$\int_s^t dW(u) = \int_s^t \frac{\partial W(u)}{\partial u} du,$$

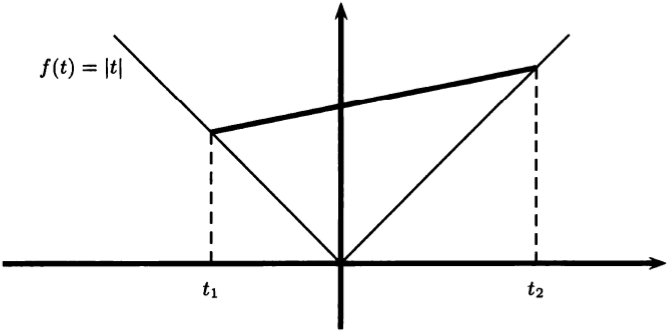
which is an ordinary (Lebesgue) integral wrt time

- This will not work for Brownian Motion because Brownian paths cannot be differentiated wrt time

Why is that?

Reason is that Brownian Motion is a **“pointy” process**, as we have observed in the earlier picture.

To illustrate: an example of a “pointy” function



The graph on the previous slide shows the absolute value function: $f(t) = |t|$:

- The derivative is always -1 for $t < 0$
- The derivative is always 1 for $t > 0$
- The derivative is undefined in 0

The paths of Brownian Motion are very "pointy" $\implies$ there is no value for t for which $\frac{d}{dt}W(t)$ is defined.

One of the consequences of this property of Brownian Motion is that the quadratic variation of a path is **not** zero!

For ordinary functions, that have a continuous derivative, quadratic variation is equal to zero (see section 3.4.2 in Shreve).

However, as said, the derivative $\frac{d}{dt}W(t)$ is not defined: quadratic variation of a Brownian Motion path up to time T appears to be $[W, W](T) = T$.

This property marks the difference between **ordinary calculus** and **stochastic calculus**.

Another consequence is that we need to give a proper meaning to the **stochastic integral**:

$$\int_s^t dW(u).$$

That will be the topic of the next knowledge clip.

Part IV

Clip 4

Previous knowledge clip:

- For a Brownian Motion process W, the derivative $\frac{d}{dt}W(t)$ is not defined
- Consequence: the **quadratic variation** of a Brownian path up to time T is not equal to zero (but equals T)
- Another consequence: we need to give proper meaning to the **stochastic integral** $\int_s^t dW(u)$ as we cannot use $\int_s^t \frac{\partial W(u)}{\partial u} du$

So, we have a so-called **stochastic integral**.

Let us consider a more general version of this integral:

$$\int_0^T \sigma(t)dW(t).$$

The question is: can we give a proper meaning to this stochastic integral?

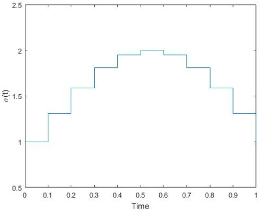
- Suppose $\sigma(t)$ is a piecewise constant function, changing value at times

$$0 = t_0 < t_1 < t_2 < \ldots < t_n = T$$

- Then we can define the integral as follows

$$Z_n(T) = \int_0^T \sigma(t)dW(t) = \sum_{i=0}^{n-1} \sigma(t_i, t_{i+1})(W(t_{i+1}) - W(t_i))$$

$Z_n(T)$ is a Normal random variable, since this a sum of (weighted) Normals.



- If $\sigma(t)$ is a **continuous function** on $[0, T]$, we approximate it by a piecewise function $\sigma_n(t)$.
- By making the step $\Delta t = t_{i+1} - t_i$ smaller and smaller we can get better and better approximation of $\sigma_n(t)$ to $\sigma(t)$.
- We define our stochastic integral as the limit

$$Z(T) = \int_0^T \sigma(t)dW(t) = \lim_{n\to\infty} \sum_{i=0}^{n-1} \sigma_n(t_i, t_{i+1})(W(t_{i+1}) - W(t_i))$$

So, $Z_n(T) \to Z(T)$ as $n \to \infty$ or $\Delta t_i \to 0$.

So, if we return to the stochastic integral we started with:

$$\int_s^t dW(u),$$

then we have times:

$$s = t_0 < t_1 < t_2 < \ldots < t_n = t,$$

and is the stochastic integral defined as follows:

$$Z_n(s,t) = \int_s^t dW(u) = \sum_{i=0}^{n-1} 1 \cdot (W(t_{i+1}) - W(t_i)) = W(t_n) - W(t_0).$$

Anything that depends on the path of a random process is itself random, i.e. the stochastic integral is a **random variable**.

The properties $\int_s^t dW(u)$ can easily be derived from the definition:

$$\mathbb{E}\left[\int_s^t dW(u)\right] = \mathbb{E}\left[W(t) - W(s)\right] = 0$$

$$\mathbb{E}\left[\left(\int_s^t dW(u)\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=0}^{n-1} 1 \cdot (W(t_{i+1}) - W(t_i))\right)^2\right] = t - s$$

In general, we can derive two important properties of stochastic integrals (where functions $g(s)$ satisfy some technical conditions):

1. $\mathbb{E}\left[\int_0^T g(s)dW(s)\right] = 0$,
2. $\mathbb{E}\left[\left(\int_0^T g(s)dW(s)\right)^2\right] = \int_0^T \mathbb{E}[g(s)^2]ds$ – **Itô isometry**.

Note that the stochastic integral $Z(T)$ is a normal variable with mean 0 and variance $\int_t^T E[g(s)^2]ds$.

The simplest way to see this is to look at our approximation $Z_n(t)$:

1. $\mathbb{E}[Z_n] = \mathbb{E}\left[\sum_{i=0}^{n-1} g_n(t_i)\left(W(t_{i+1}) - W(t_i)\right)\right] = 0,$
2. $\mathbb{E}[Z_n^2] = \mathbb{E}\left[\sum_{i=0}^{n-1} g_n(t_i)\left(W(t_{i+1}) - W(t_i)\right)\right]^2 = \sum_{i=0}^{n-1} \mathbb{E}\left[g_n^2(t_i)\right](t_{i+1} - t_i).$

We use Brownian Motions to build more interesting processes:

$$dX(t) = \mu(\cdot)dt + \sigma(\cdot)dW(t),$$

which is short-hand notation for the more “proper” mathematical expression

$$X(T) - X(0) = \int_0^T \mu(t)dt + \int_0^T \sigma(t)dW(t)$$

We call these stochastic processes **diffusion**.
$\mu(\cdot)dt$ is called the **drift** term.
$\sigma(\cdot)dW(t)$ is called the **diffusion** term.
Respectively,
μ is a **drift**.
σ is called **volatility**.

- $\mu(\cdot)$ and $\sigma(\cdot)$ cannot be some weird functions. They have to be deterministic functions.
- But, they can depend on $X(t)$ and on time t.
- What is important is that $\mu(\cdot)$ and $\sigma(\cdot)$ should be **adapted** to the **filtration**. In simple words, all variables, on which the drift and diffusion depended on, should be "known" at time t.
- Let's give an example:

$$dX(t) = \sin(X(t)^2)dt + \cos(t)dW(t) \quad \text{is ok}$$
$$dX(t) = \sin(X(t+1)^2)dt + \cos(t)dW(t) \quad \text{is not ok}$$

The increment of $X(t)$ at time t would depend on some future values $X(t+1)$, which are not known at time t, still uncertain. We will not study this kind of processes. Remember, $dW(t)$ is a forward-looking and is not known yet.

Part V

Summary

This course is about finding the **no-arbitrage price** of **derivatives** contracts.

No-arbitrage pricing means that prices are determined in such a way that there is **no free lunch** in the market.

Björk defines an **arbitrage portfolio** as a portfolio h with the properties:

$$V_0^h = 0$$
$$V_1^h > 0 \text{ with probability } 1.$$

An arbitrage portfolio is thus a deterministic money making machine.

One way of deriving the no-arbitrage price is by means of employing the risk-neutral valuation method:

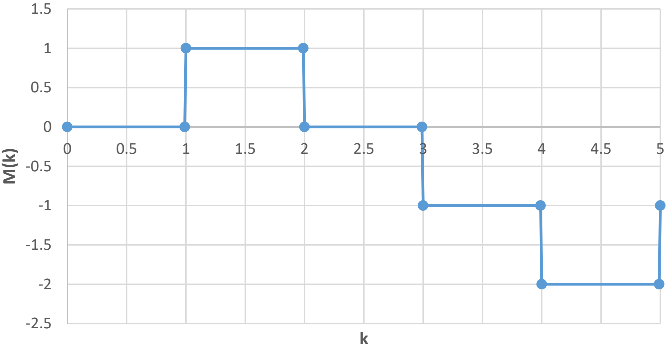
$$V_t = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left(V_T|\mathcal{F}_t\right).$$

In order to be able to evaluate the conditional expecation (under $\mathbb{Q}$), we need to specifiy a model for the underlying value S of the derivative contract.

In the first week we used the **discrete time binomial tree model** as a model for S.

This week we introduced the continuous time stochastic process W: **Brownian Motion**.

We started with looking at the **symmetric random walk** process M:



To approximate a **Brownian Motion process**, we

- speed up time, i.e. we are going to toss the coin more often
- scale down the step size, i.e. the step size will be smaller than 1 (in absolute terms)

More precisely, we fix a positive integer n and define the **scaled symmetric random walk** as follows:

$$W^n(t) = \frac{1}{\sqrt{n}} M_{nt}.$$

Example:

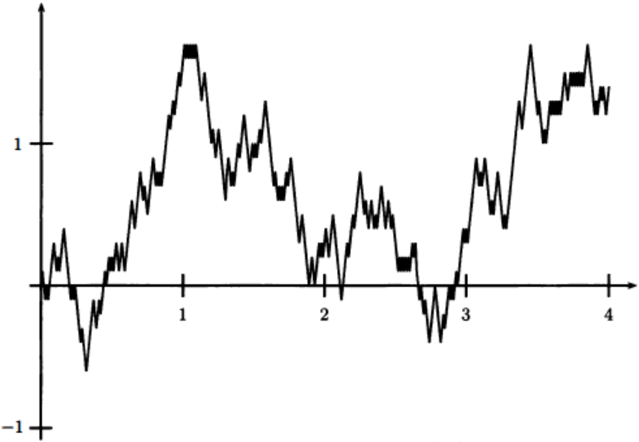


Fig. 3.2.2. A sample path of $W^{(100)}$.

What happens if $n \to \infty$? Then we get the continuous time Brownian Motion process W:

$$W(t) = \lim_{n\to\infty} W^n(t).$$

Properties:

- $W(0) = 0$
- Increments are independent
- Increments are normally distributed
- Increments have expectation equal to zero
- Increments have variance equal to the time span of the increment, e.g. $\mathrm{Var}(W(3)) = \mathrm{Var}(W(3) - W(0)) = 3$

For the scaled symmetric random walk we could say:

$$\Delta W^n_{t_i} = W^n_{t_{i+1}} - W^n_{t_i}.$$

In the **continuous time framework**, we write:

$$dW(t).$$

Simply stated, Δ *becomes very small, infinitely small (infinitesimal)*.

Like Δ, d is a *forward looking operator*:

$$dW(t) = W(t+dt) - W(t), \quad \text{as is} \quad \Delta W^n_{t_i} = W^n(t_i+\Delta) - W^n(t_i)$$

So, $dW(t)$ is the increment of Brownian Motion over an infinitesimal timestep.

Hence, one way of specifying a continuous time process for stock price S could be by means of the following **stochastic differential equation**:

$$dS(t) = dW(t),$$

which is shorthand notation for:

$$S(T) - S(0) = \int_0^T dW(s).$$

In words: the increment in the stock price S between $t = 0$ and $t = T$ is given by summing the instantaneous changes, i.e. the increments over infinitesimal time steps, of Brownian Motion between $t = 0$ and $t = T$.

The integral:

$$\int_0^T g(s)dW(s),$$

is called a **stochastic integral**.

Properties:

1. $\mathbb{E}\left[\int_0^T g(s)dW(s)\right] = 0,$
2. $\mathbb{E}\left[\left(\int_0^T g(s)dW(s)\right)^2\right] = \int_0^T \mathbb{E}[g(s)^2]ds$ – **Itô isometry**.

Slide deck week 4

Mark-Jan Boes

September, 2020

Part I

Clip 1

Last week we talked about:

- Brownian Motion: a continuous stochastic process W
- Stochastic integration

For $s < t$ we can write the increment in Brownian Motion between s and t as:

$$\underbrace{W(t) - W(s)}_{\text{increment}} = \underbrace{\int_s^t dW(u)}_{\text{stochastic integral}} \quad .$$

In words: the increment in W between s and t is given by summing the instantaneous changes, i.e. the increments over infinitesimal time steps, of Brownian Motion between s and t.

A stochastic process W is called a **Brownian motion** if the following conditions hold:

1. $W(0) = 0$
2. $W(u) - W(t)$ and $W(s) - W(r)$ are **independent stochastic variables for non-overlapping intervals**, $r < s < t < u$.
3. $W(t) - W(s)$ is **normally distributed** with mean 0 and variance $t - s$

$$\mathbb{E}[W(t) - W(s)] = 0, \qquad Var[W(t) - W(s)] = t - s$$

4. W has continuous trajectories.

Let $W(t)$ be a standard Brownian motion.

- **The Brownian motion is a martingale process:**

$$\begin{aligned} \mathbb{E}\left[W(t)|\,\mathcal{F}_s\right] &= \mathbb{E}\left[W(t) - W(s)|\,\mathcal{F}_s\right] + \mathbb{E}\left[W(s)|\,\mathcal{F}_s\right] \\ &= 0 + W(s) \\ \text{or} \qquad \mathbb{E}\left[W(t)|\,\mathcal{F}_s\right] &= W(s) \qquad \text{for} \qquad 0 \leq s \leq t. \end{aligned}$$

this follows from the independence of increments.

Variance of the Brownian motion is not a martingale

$$\mathbb{E}\left[W^2(t) \middle| \mathcal{F}_s \right] \geq W^2(s) \qquad \text{for} \qquad 0 \leq s \leq t.$$

But we can find the compensation

$$\begin{aligned} \mathbb{E}\left[W^2(t) \middle| \mathcal{F}_s \right] = \\ &= \mathbb{E}\left[(W(t) - W(s))^2 + 2\,(W(t) - W(s)) \cdot W(s) + W^2(s) \middle| \mathcal{F}_s \right] \\ &= (t - s) + 0 + W^2(s) \qquad \text{for} \qquad 0 \leq s \leq t, \end{aligned}$$

which shows that the process

$$W^2(t) - t, \qquad t \geq 0$$

is a martingale.

Brownian motion is a **Markov process**:

- To see this:

$$\mathbb{E}\left[g(W(t))\middle|\,\mathcal{F}_s\right] = \mathbb{E}\left[g(W(t))\middle|\,W(s)\right] \qquad \text{for} \qquad 0 \leq s \leq t,$$

for every bounded continuous function $g(\cdot)$.

In other words, the future path of Brownian Motion depends on the current value $W(s)$, but not on the values before it.

This follows from the fact that $W(t) - W(s)$ is independent of $\mathcal{F}_s$.

Covariance of two random variables $W(s)$ and $W(u)$ for any two times $0 \leq s < u$ is

$$\begin{aligned}\mathbb{E}\left[W(s)W(u)\right] &= \mathbb{E}\left[W(s)\left(W(u) - W(s)\right) + W^2(s)\right] \\ &= \mathbb{E}\left[W(s)\right]\mathbb{E}\left[W(u) - W(s)\right] + \mathbb{E}\left[W^2(s)\right] \\ &= 0 + Var\left[W(s)\right] = s\end{aligned}$$

or generally:

$$\mathbb{E}\left[W(s)W(u)\right] = s \wedge u = \min(s, u)$$

and correlation

$$\rho = \frac{\mathbb{E}\left[W(s)W(u)\right]}{\sqrt{Var[W(s)]Var[W(u)]}} = \frac{s}{\sqrt{s \cdot u}} = \sqrt{\frac{s}{u}}$$

Last week we looked at the quadratic variation of the **symmetric random walk**.

$$[M, M]_k = \sum_{j=1}^{k} (M_j - M_{j-1})^2 = \sum_{j=1}^{k} 1 = k.$$

In words: we simply square the steps a process takes (i.e. the increments) along a specific path.

What can we say about **quadratic variation for Brownian Motion**? And why is it relevant?

For Brownian Motion there is no natural step size as it is a **continuous process with infinitesimal time steps**.

A logical approach would be to, for a given $T > 0$, choose a step size $\frac{T}{n}$ for some large n, and compute the quadratic variation up to time T with this step size:

$$[W, W]_T = \sum_{j=0}^{n-1} \left[W\left(\frac{(j+1)T}{n}\right) - W\left(\frac{jT}{n}\right)\right]^2.$$

And then take the limit $n \to \infty$.

Let us first look at **First-Order Variation**:

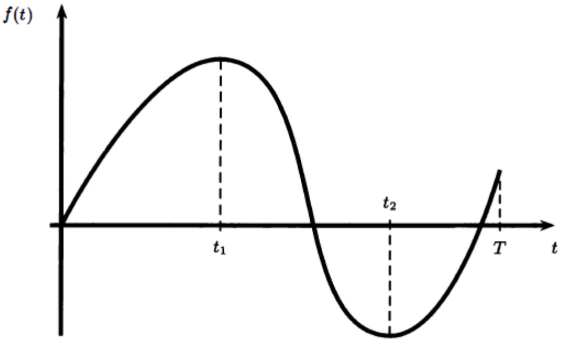


Fig. 3.4.1. Computing the first-order variation.

$$FV_T(f) = (f(t_1) - f(0)) - (f(t_2) - f(t_1)) + (f(T) - f(t_2)).$$

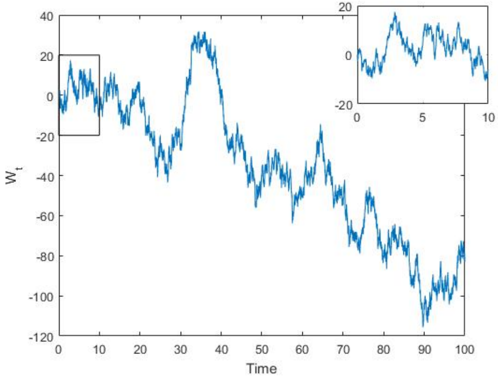
Then it is easy to see that the formula for first order variation can be rewritten as:

$$FV_T(f) = \int_0^T |f'(t)| dt.$$

Summing the absolute changes in the function should give the amount of up and down oscillation.

We need to have a function that behaves nicely, i.e. it should have a continuous derivative.

What do we know about Brownian Motion?



It is 'erratic', 'pointy', 'spiky': there is no value for t for which $\frac{d}{dt}W(t)$ is defined.

Let's move to quadratic variation: choose a partition $\Pi = t_0, t_1, \ldots, t_n$ on $[0, T]$ which is a set of times

$$0 = t_0 < t_1 < \ldots < t_n = T.$$

The maximum step size is:

$$||\Pi|| = \max_{j=0,\ldots,n-1} (t_{j+1} - t_j).$$

Quadratic variation of a function f is then defined as:

$$[f, f]_T = \lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|^2.$$

Using the mean-value theorem:

$$\sum_{j=0}^{n-1} |f(t_{j+1} - f(t_j)|^2 = \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j)^2$$
$$\leq ||\Pi|| \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j).$$

Hence,

$$[f,f]_T \leq \lim_{||\Pi||\to 0} ||\Pi|| \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j)$$
$$= \lim_{||\Pi||\to 0} ||\Pi|| \lim_{||\Pi||\to 0} \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j) = 0.$$

Most functions have continuous derivatives and hence their quadratic variation is zero.

For this reason quadratic variation is irrelevant in ordinary calculus.

As was already mentioned last week, the paths of Brownian Motion cannot be differentiated with respect to the time variable.

So, we **cannot** conclude that the quadratic variation of Brownian Motion is zero. In fact:

$$[W, W]_T = T.$$

The Brownian Motion is the continuous time version of the Random Walk.

As a consequence, Brownian Motion is **not** a smooth stochastic process, leading to the following properties of a Brownian Motion up to time T:

- The First-Order Variation in $[0, T]$ is equal to ∞ for all $T > 0$
- The Quadratic Variation in $[0, T]$ is equal to T for all $T > 0$

The second property has as an implication that stochastic calculus differs from ordinary calculus, as we will see later this week when we discuss Itô's Lemma.

Part II

Clip 2

Last week we talked about:

- Brownian Motion: a continuous stochastic process W
- Stochastic integration

For $s < t$ we can write the increment in Brownian Motion between s and t as:

$$W(t) - W(s) = \int_s^t dW(u).$$

In words: the increment in W between s and t is given by summing the instantaneous changes, i.e. the increments over infinitesimal time steps, of Brownian Motion between s and t.

A stochastic process W is called a **Brownian motion** if the following conditions hold:

1. $W(0) = 0$
2. $W(u) - W(t)$ and $W(s) - W(r)$ are **independent stochastic variables for non-overlapping intervals**, $r < s < t < u$.
3. $W(t) - W(s)$ is **normally distributed** with mean 0 and variance $t - s$

$$\mathbb{E}[W(t) - W(s)] = 0, \qquad Var[W(t) - W(s)] = t - s$$

4. W has continuous trajectories.

- **Conditional distribution**: From the definition, the increments are normal random variables $W_T - W_t \sim N(0, T-t)$. Thus the distribution of the random variable W_T **given** $W_t = x$ is

$$W_T \sim N(x, T-t) \qquad \text{condition on} \quad W_t = x$$

or in terms of probability density function:

$$p(t,x;T,y)dy := \mathbb{P}\left[W_T \in [y, y+dy] \middle| W_t = x \right] = \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{(y-x)^2}{2(T-t)}} dy$$

- **Cumulative distribution function**

$$\begin{aligned}\mathbb{P}\left[W_T < z \middle| W_t = x \right] &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{(y-x)^2}{2(T-t)}} dy = N\left(z; x, T-t\right) \\ &= N\left(\frac{z-x}{\sqrt{T-t}}\right),\end{aligned}$$

where $N(x)$ is a CDF for standard normal distribution $N(0,1)$.

- **Conditional expectation** of function $g(\cdot)$ of Brownian Motion can thus be defined as

$$V(t,x) = \mathbb{E}\left[g\left(W_T\right)\middle|\, \mathcal{F}_t\right] = \mathbb{E}\left[g\left(W_T\right)\middle|\, W_t = x\right] = \int_{-\infty}^{\infty} g(y)p(t,x;T,y)dy.$$

- Example: Conditional expectation of exponent

$$V(t,x) = \mathbb{E}\left[e^{W_T}\middle|\, W_t = x\right] = \int_{-\infty}^{\infty} e^y \cdot \frac{e^{-\frac{(y-x)^2}{2(T-t)}}}{\sqrt{2\pi(T-t)}}dy$$

$$= e^x \int_{-\infty}^{\infty} \frac{e^{z-\frac{z^2}{2(T-t)}}}{\sqrt{2\pi(T-t)}}dz$$

$$= e^x \int_{-\infty}^{\infty} \frac{e^{-\frac{(z-(T-t))^2-(T-t)^2}{2(T-t)}}}{\sqrt{2\pi(T-t)}}dz = e^{x+\frac{T-t}{2}} = e^{W_t+\frac{T-t}{2}},$$

which also shows that $e^{W_t - t/2}$ is a martingale: the **exponential martingale**.

- Another example: **Expectation of power function**

$$\mathbb{E}\left[W_t^m\right], \qquad m = 1, 2, 3, \ldots$$

- Let us the case
  - Odd: $m = 2n + 1 \qquad n = 0, 1, 2, 3, \ldots$
- In case of odd powers

$$\mathbb{E}_0[W_t^{2n+1}] = 0$$

Indeed, in this case the integrand is an antisymmetric function and thus

$$\begin{aligned}
\mathbb{E}[W_t^{2n+1}] &= \int_{-\infty}^{\infty} y^{2n+1} \cdot \frac{e^{-\frac{y^2}{2t}}}{\sqrt{2\pi t}} dy \\
&= \int_{-\infty}^{0} y^{2n+1} \cdot \frac{e^{-\frac{y^2}{2t}}}{\sqrt{2\pi t}} dy + \int_{0}^{\infty} y^{2n+1} \cdot \frac{e^{-\frac{y^2}{2t}}}{\sqrt{2\pi t}} dy = (z = -y) = \\
&= -\int_{0}^{\infty} z^{2n+1} \cdot \frac{e^{-\frac{z^2}{2t}}}{\sqrt{2\pi t}} dz + \int_{0}^{\infty} y^{2n+1} \cdot \frac{e^{-\frac{y^2}{2t}}}{\sqrt{2\pi t}} dy = 0
\end{aligned}$$

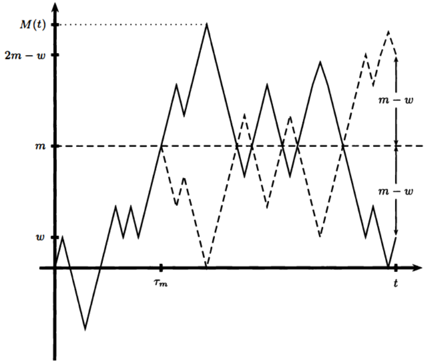
Let W_t be a standard Brownian motion. Then the following processes are also Brownian motion

- $-W_t$ (symmetry),
- $\frac{1}{\sqrt{c}} W_{ct}$ (scaling),

For proof, check the properties in the definition of Brownian motion.

- Consider a Brownian Motion path starting at zero and some level $m > 0$.
- Note that for every path that reaches level m before time t and ends at level w below m at time t there is a "reflected path" that ends at $2m - w$.
- This leads to the reflection equality

$$\mathbb{P}\{\tau_m \leq t, W(t) \leq w\}$$
$$= \mathbb{P}\{W(t) \geq 2m - w\}, \quad w \leq m, m > 0.$$



- Suppose we want to find probability of passing level m before time t. Let us call this random variable $\tau_m = \inf(t : W(t) \geq m)$.
- From the reflection principle we have

$$\mathbb{P}\{\tau_m \leq t, W(t) \leq w\} = \mathbb{P}\{W(t) \geq 2m - w\}, \quad w \leq m, m > 0.$$

- Lets first compute the probability to cross the level m and finish at level $w \leq m$. In the equation above we set $w = m$

$$\mathbb{P}\{\tau_m \leq t, W(t) \leq m\} = \mathbb{P}\{W(t) \geq m\}.$$

- Second, if $W(t) \geq m$, then $\tau_m \leq t$:

$$\mathbb{P}\{\tau_m \leq t, W(t) \geq m\} = \mathbb{P}\{W(t) \geq m\}.$$

These two equations give

$$\begin{aligned}\mathbb{P}\{\tau_m \leq t\} &= \mathbb{P}\{\tau_m \leq t, W(t) \leq m\} + \mathbb{P}\{\tau_m \leq t, W(t) \geq m\} \\ &= 2\mathbb{P}\{W(t) \geq m\} = \frac{2}{\sqrt{2\pi t}} \int_m^\infty e^{-\frac{x^2}{2t}}\, dx.\end{aligned}$$

Part III

Clip 3

Last week we talked about:

- Brownian Motion: a continuous stochastic process W
- Stochastic integration

For $s < t$ we can write the increment in Brownian Motion between s and t as:

$$W(t) - W(s) = \int_s^t dW(u).$$

In words: the increment in W between s and t is given by summing the instantaneous changes, i.e. the increments over infinitesimal time steps, of Brownian Motion between s and t.

A stochastic process W is called a **Brownian motion** if the following conditions hold:

1. $W(0) = 0$
2. $W(u) - W(t)$ and $W(s) - W(r)$ are **independent stochastic variables for non-overlapping intervals**, $r < s < t < u$.
3. $W(t) - W(s)$ is **normally distributed** with mean 0 and variance $t - s$

$$\mathbb{E}[W(t) - W(s)] = 0, \qquad Var[W(t) - W(s)] = t - s$$

4. W has continuous trajectories.

Next question is: suppose we have a function $F(t) = F(t, X(t))$.

What is the process for $F(t)$? What is the differential dF?

Consider continuous function $F(t,x)$ of two **non-stochastic** variables t and x. Using a Taylor series expansion we can write $F(t+\Delta t, x+\Delta x)$ around $F(t,x)$ as:

$$F(t+\Delta t, x+\Delta x) \approx F(t,x) + \frac{\partial F}{\partial t}\Delta t + \frac{\partial F}{\partial x}\Delta x + \frac{1}{2}\frac{\partial^2 F}{\partial t^2}\Delta t^2 + \frac{\partial^2 F}{\partial t \partial x}\Delta t \Delta x + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\Delta x^2 + \dots$$

The **infinitesimal increment** dF is

$$dF(t,x) = \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial x}dx + o(dt, dx)$$

the term $o(dt,dx)$ contains all other terms of the Taylor expansion that are smaller than dt and dx, like

$$o(dt,dx) = \frac{1}{2}\frac{\partial^2 F}{\partial t^2}dt^2 + \frac{\partial^2 F}{\partial t \partial x}dtdx + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}dx^2 + \dots$$

In **ordinary calculus**, we **ignore** all the terms smaller than dt and dx: only first order effects are relevant.

If x is a function of time and some other variable v, such that $x = x(t, v)$, we would have for x and $F(t, x)$

$$dx(t, v) = \frac{\partial x}{\partial t}\, dt + \frac{\partial x}{\partial v}\, dv = a(t, v)dt + b(t, v)dv$$
$$dF(t, x) = \left(\frac{\partial F}{\partial t} + a \frac{\partial F}{\partial x} \right) dt + b \frac{\partial F}{\partial x}\, dv$$

If $X(t)$ is the following **stochastic process** (in differential form):

$$dX(t) = \mu dt + \sigma dW(t),$$

and $F = F(t, X)$, then in case of ordinary calculus, and $F(t, X)$ an ordinary function of two variables, we would stop by the linear terms

$$dF(t, X) = \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial X} dX \dots$$

But this is **stochastic calculus!** We need the second order term for stochastic X:

$$dF(t, X) = \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial X} dX + \boldsymbol{\frac{1}{2} \frac{\partial^2 F}{\partial X^2} dX^2} + \dots$$

If we substitute $dX = \mu dt + \sigma dW$, we get

$$dF = \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial X}(\mu dt + \sigma dW) + \frac{1}{2}\frac{\partial^2 F}{\partial X^2}(\mu^2 dt^2 + 2\mu\sigma \cdot dtdW + \sigma^2 dW^2)$$

We set $\boldsymbol{dt^2} = 0$, $\boldsymbol{dtdW} = 0$, and $\boldsymbol{dW^2 = dt}$ to get the Itô's formula:

$$dF = \frac{\partial F}{\partial t}dt + \mu\frac{\partial F}{\partial X}dt + \sigma\frac{\partial F}{\partial X}dW + \frac{1}{2}\frac{\partial^2 F}{\partial X^2}\sigma^2 dt$$

Why should we set $\boldsymbol{dW^2 = dt}$?

- Intuition 1: Let's first take an expectation from both sides in our Taylor expansion

$$dF = \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial X}(\mu dt + \sigma dW) + \frac{1}{2}\frac{\partial^2 F}{\partial X^2}(\mu^2 dt^2 + 2\mu\sigma \cdot dt dW + \sigma^2 dW^2)$$

After taking expectation all 'diffusion' terms disappear, but one term

$$\mathbb{E}_t[dF] = \mathbb{E}_t\left[\frac{\partial F}{\partial t}\right] dt + \mathbb{E}_t\left[\frac{\partial F}{\partial X}\mu\right] dt + \mathbb{E}_t\left[\frac{1}{2}\frac{\partial^2 F}{\partial X^2}\sigma^2\right] \underbrace{\mathbb{E}_t[dW^2]}_{dt}$$

The drift has an additional term (all dW give zero)

$$\frac{1}{2}\sigma^2 \frac{\partial^2 F}{\partial X^2} \cdot dt$$

Intuition 2: Let's divide the interval $[0, t]$ into n equal subintervals $\left[k\frac{t}{n}, (k+1)\frac{t}{n}\right]$, $k = 0, 1, \ldots, n-1$. And consider the sum

$$S_n = \sum_{i=1}^{n} \left[W\left(i\frac{t}{n}\right) - W\left((i-1)\frac{t}{n}\right)\right]^2$$

The expectation and the variance of the sum are

$$E[S_n] = \sum_{i=1}^{n} E\left[W\left(i\frac{t}{n}\right) - W\left((i-1)\frac{t}{n}\right)\right]^2 = \sum_{i=1}^{n} \left(i\frac{t}{n} - (i-1)\frac{t}{n}\right) = t$$

$$Var[S_n] = \sum_{i=1}^{n} Var\left[W\left(i\frac{t}{n}\right) - W\left((i-1)\frac{t}{n}\right)\right]^2 = \sum_{i=1}^{n} 2\left[\frac{t^2}{n^2}\right] = 2\frac{t^2}{n}$$

That is $E[S_n] \to t$ and $Var[S_n] \to 0$ as $n \to \infty$.

In other words, S_n converges to deterministic function t:

$$\sum_{i=1}^{n} \left[W\left(i\frac{t}{n}\right) - W\left((i-1)\frac{t}{n}\right) \right]^2 \to \int_0^t dW^2(u) = t, \quad \text{or} \quad dW^2 = dt.$$

Consider the following stochastic process with constant μ and σ

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t)$$

This process is **Geometric Brownian motion**.

Define $Y(t) := f(X) = \ln X(t)$. Question: what is $dY(t)$?

We use Itô's formula:

$$\frac{\partial f}{\partial X} = \frac{1}{X}, \qquad \frac{\partial^2 f}{\partial X^2} = -\frac{1}{X^2}, \qquad \frac{\partial f}{\partial t} = 0.$$

Using Itô formula

$$dY(t) = \left(\frac{1}{X}\mu X(t) - \frac{1}{2}\frac{1}{X^2}\sigma^2 X^2(t)\right)dt + \frac{1}{X}\sigma X(t)dW(t)$$
$$= \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW(t)$$

The equation becomes much simpler and can be solved analytically:

$$Y(T) = Y(0) + \left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma[W(T) - W(0)].$$

Part IV

Clip 4

In the previous knowledge clip, we introduced **Itô's formula for Brownian Motion**:

Let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x)$, $f_x(t,x)$, and $f_{xx}(t,x)$ are defined and continuous, and let $W(t)$ be a Brownian Motion. Then for every $T > 0$,

$$f(T, W(T)) = f(0, W(0)) + \int_0^T f_t(t, W(t))dt + \int_0^T f_x(t, W(t))dW(t) + \boldsymbol{\frac{1}{2}\int_0^T f_{xx}(t, W(t))dt}.$$

Or (in differential form):

$$df(t) = f_t(t, W(t))dt + f_x(t, W(t))dW(t) + \boldsymbol{\frac{1}{2} f_{xx}(t, W(t))dt}.$$

Compared to ordinary calculus, we observe an additional term which we call the Itô term:

$$\frac{1}{2} f_{xx}(t, W(t))dt.$$

This additional term apears because quadratic variation of a Brownian path is non-zero.

In the previous knowledge clip we argued that the quadratic variation of a Brownian path up to time T equals T regardless of the path along which we are doing our calculations.

Consequence:

$$d[W, W]_t = [W, W]_{t+dt} - [W, W]_t = t + dt - t = dt.$$

Consider the following stochastic integral:

$$Y(t) = \int_0^t W(s)\, dW(s).$$

Can we use Itô's lemma to solve this stochastic integral?

If W were not stochastic, we would have

$$\tilde{Y}(t) = \frac{1}{2} W^2(t).$$

Let's apply Itô's lemma to this function of a Brownian Motion:

$$d\tilde{Y}(t) = W(t) dW(t) + \frac{1}{2} dt.$$

We can write this stochastic differential equaion in integral form:

$$\tilde{Y}(t) = \tilde{Y}(0) + \int_0^t W(s)\, dW(s) + \frac{1}{2}t.$$

Given that $Y(0) = 0$ we can now easily see that:

$$Y(t) = \tilde{Y}(t) - \frac{1}{2}t = \frac{1}{2}W^2(t) - \frac{1}{2}t.$$

Compared to ordinary calculus, the Itô term makes the difference.

Next step is to consider functions that depend on two stochastic processes: both processes can depend on more than one Brownian Motion.

Let $f(t, x, y)$ be a function for which the partial derivatives $f_t, f_x, f_y, f_{xx}, f_{xy}$, and f_{yy} are defined and continuous, and let $X(t)$ and $Y(t)$ be stochastic processes with Brownian Motion. Then,

$$df(t, X, Y) = f_t dt + f_x dX + f_y dY + \frac{1}{2} f_{xx} dX dX + f_{xy} dX dY + \frac{1}{2} f_{yy} dY dY.$$

Suppose we have M stochastic processes and we want to model them with N Wiener processes. We can write the processes in vector form:

$$d\vec{X}(t) = \vec{\mu}dt + \hat{\sigma}d\vec{W}(t), \qquad \text{or per element}$$
$$dX_i(t) = \mu_i dt + \sigma_{i1}dW_1 + \sigma_{i2}dW_2 + \ldots \sigma_{iN}dW_N$$

Here, $\vec{X}$, $\vec{\mu}$, and $\vec{W}$ are vectors, and $\hat{\sigma} = ||\sigma_{ij}||$ is a matrix.

If $F = F(\vec{X}, t)$ then we can do the very same Taylor expansion

$$dF = \frac{\partial F}{\partial t}dt + \sum_i \frac{\partial F}{\partial X_i}dX_i + \frac{1}{2}\sum_{i,j}\frac{\partial^2 F}{\partial X_i \partial X_j}dX_i dX_j.$$

We need to make assumptions about how the increments of different process are correlated. For instance, we could assume independence in increments:

$$\mathbb{E}_t[dW_i(t) \cdot dW_j(t)] = 0.$$

However, in a more general case they are correlated:

$$\mathbb{E}_t[dW_i(t) \cdot dW_j(t)] = \rho_{ij} dt.$$

Suppose that the SDEs of stochastic processes X and Y are given by:

$$dX = \alpha X dt - Y dW, \quad X(0) = x_0,$$
$$dY = \alpha Y dt + X dW, \quad Y(0) = y_0,$$

where the initial values x_0, y_0 are deterministic constants.

What can we say about the process R defined by $R(t) = X^2(t) + Y^2(t)$?

We apply the Itô formula to $R = R(X, Y)$:

$$\begin{aligned}
dR &= \partial_X R dX + \partial_Y R dY + \frac{1}{2}\partial_{XX} R dX^2 + \frac{1}{2}\partial_{YY} R dY^2 \\
&= 2XdX + 2YdY + dX^2 + dY^2 \\
&= 2\alpha X^2 dt - 2XYdW + 2\alpha Y^2 dt + 2XYdW + Y^2 dt + X^2 dt \\
&= (2\alpha + 1)(X^2 + Y^2)dt = (2\alpha + 1)Rdt.
\end{aligned}$$

The absence of the diffusion term shows that the process $R(t)$ is deterministic.

Part V

Clip 5

Last week we talked about:

- Brownian Motion: a continuous stochastic process W
- Stochastic integration

We introduced to stochastic integral as:

$$\int_0^t dW(u) = W(t) - W(0).$$

Using the properties of Brownian Motion:

$$\int_0^t dW(u) \sim N(0, t).$$

We could calculate the variance of the stochastic integral as:

$$Var\left(\int_0^t dW(u)\right) = Var(W(t) - W(0)) = t.$$

But also as:

$$\begin{aligned} Var\left(\int_0^t dW(u)\right) &= \mathbb{E}\left[\left(\int_0^t dW(u)\right)^2\right] \\ &= \mathbb{E}\left[\left(\int_0^t dW(u)\right)\left(\int_0^t dW(u)\right)\right] \\ &= \mathbb{E}\left[\int_0^t dW^2(u)\right] = \mathbb{E}[t] = t. \end{aligned}$$

We looked at more general stochastic integrals as well:

$$\int_0^t g(s)dW(s).$$

Then:

$$\mathbb{E}\left(\int_0^t g(s)dW(s)\right) = \mathbb{E}\left(\int_0^t \mathbb{E}_s(g(s)dW(s))\right)$$
$$= \mathbb{E}\left(\int_0^t g(s)\mathbb{E}_s(dW(s))\right) = 0.$$

And:

$$Var\left(\int_0^t g(s)dW(s)\right) = \int_0^t \mathbb{E}[g(s)^2]ds.$$

But what can we say about:

$$\int_s^t g(u)dW(u),$$

with $0 \leq s < t$.

We need to rely on conditional expecation:

$$\mathbb{E}_s[Y] = \mathbb{E}[\,Y|\,\mathcal{F}_s],$$

where $\mathcal{F}_s$ represents the information available up to time s

In the same way as before, we can derive

$$\mathbb{E}\left[\int_s^t g(u)dW(u)\middle|\mathcal{F}_s\right] = \mathbb{E}_s\left[\int_s^t g(u)dW(u)\right] = 0,$$

and

$$\mathbb{E}_s\left[\left(\int_s^t g(u)dW(u)\right)^2\right] = \int_s^t \mathbb{E}_s[g(u)^2]du.$$

Recall, that a stochastic process X_t is called an $\mathcal{F}_t$-**martingale** if it satisfies some technical conditions and

$$\mathbb{E}_s[X(t)] = \mathbb{E}_s[X(t) | \mathcal{F}_s] = X(s)$$

for $s \leq t$.

"The best estimate for X_t is its current value".

We know from the properties of stochastic integrals:

$$\mathbb{E}\left[\int_s^t g(u)dW(u)\middle|\mathcal{F}_s\right] = 0$$

A **stochastic process** Z defined as

$$Z(t) = \int_0^t g(u)dW(u)$$

is an $\mathcal{F}_t$-martingale.

Proof: Split the integral in two: $Z(t) = Z(s) + \int_s^t g(u)dW(u)$
then $\mathbb{E}_s[Z(t)] = \mathbb{E}_s[Z(s)] + \mathbb{E}_s[\int_s^t g(u)dW(u)] = Z(s)$.

Why is all of this relevant?

Well, we could try to model stock prices as a continuous time process, for instance as:

$$dS(t) = \sigma dW(t).$$

Or in integral notation:

$$S(t) - S(0) = \sigma \int_0^t dW(u).$$

Econometricians will object immediately against this proposal.

The reason is that in discrete time the process can be rewritten as:

$$S_{t+h} = S_t + \varepsilon_{t:t+h},$$

or:

$$S_{t+h} = S_0 + \sum_{i=1}^{n} \varepsilon_{t_{i-1}:t_i},$$

with:

$$0 = t_0 < t_1 < ... < t_n = t + h,$$

the time steps equally large and all ε i.i.d. normal with expectation 0 and variance σ^2 .

The unconditional variance of S_{t+h} is given by:

$$\mathbb{E}(S_{t+h}^2|\mathcal{F}_0) - (\mathbb{E}(S_{t+h}|\mathcal{F}_0))^2 = \mathbb{E}(S_{t+h}^2|\mathcal{F}_0).$$

Using the properties of ε_t we get:

$$\mathbb{E}(S_{t+h}^2|\mathcal{F}_0) = \sigma^2(t+h).$$

This shows that unconditional variance of S_{t+h} depends on time and therefore the process for S is non-stationary.

Economists also have a problem with the model because they know from their academic research that the risk in equity exposure is compensated for, i.e. there is a so-called equity risk premium.

The expectation of tomorrow's stock price given today's stock price is therefore positive, meaning that the process for S cannot be a martingale.

This could be fixed by including a drift term:

$$dS_t = \mu dt + \sigma dW_t.$$

Or in discrete time series notation:

$$S_{t+h} = S_t + \mu h + \varepsilon_{t:t+h}.$$

However, this does not solve the problem of non-stationarity.

Another issues of this model for stock prices:

- S can become negative

So what we could do is to reformulate the model as follows (in discrete time):

$$\begin{aligned} S_{t+h} &= S_t(1 + \mu h + \varepsilon_{t:t+h}) \\ &= S_t + \mu S_t h + S_t \varepsilon_{t:t+h}. \end{aligned}$$

Or in terms of simple net returns:

$$\frac{S_{t+h} - S_t}{S_t} = \mu h + \varepsilon_{t:t+h}.$$

This is a stationary process for simple net returns, but (still) the issue is that returns smaller than 100% are possible according to this model.

This problem is solved by converting the model to continuous time:

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t).$$

In words: the instantaneous return on S is normally distributed with mean μ and standard deviation σ.

Practically, we cannot do anything with an instantaneous return: we need to convert to discrete time.

The point here is that the discrete time distribution of simple (gross) returns implied by the stochastic differential equation on the previous page is NOT normally distributed but lognormally distributed.

This week we have learned how to apply Itô's Lemma in order to be able to derive that:

$$d \log S(t) = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW(t).$$

Or in integral form:

$$\log S(t) = \log S(0) + \int_0^t \mu du + \int_0^t \sigma dW(u).$$

It also holds that:

$$\log S(t+h) = \log S(0) + \int_0^{t+h} \mu du + \int_0^{t+h} \sigma dW(u).$$

And therefore:

$$\log S(t+h) = \log S(t) + \int_t^{t+h} \mu du + \int_t^{t+h} \sigma dW(u).$$

Using the properties of the stochastic integral earlier this knowledge clip we have:

$$\log \left(\frac{S(t+h)}{S(t)} \middle| \mathcal{F}_t \right) \sim N(\mu h, \sigma^2 h).$$

Part VI

Summary

This course is about finding the **no-arbitrage price** of **derivatives** contracts.

No-arbitrage pricing means that prices are determined in such a way that there is **no free lunch** in the market.

Björk defines an **arbitrage portfolio** as a portfolio h with the properties:

$$V_0^h = 0$$
$$V_1^h > 0 \text{ with probability } 1.$$

An arbitrage portfolio is thus a deterministic money making machine.

One way of deriving the no-arbitrage price is by means of employing the risk-neutral valuation method:

$$V_t = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left(V_T|\mathcal{F}_t\right).$$

In order to be able to evaluate the conditional expecation (under $\mathbb{Q}$), we need to specify a model for the underlying value S of the derivative contract.

In the first week we used the **discrete time binomial tree model** as a model for S.

Last week and this week we talked about the continuous time stochastic process W: **Brownian Motion**.

The distribution of an increment of Brownian on time inteval [s, t] is given by:

$$W(t) - W(s) \sim N(0, t - s).$$

We often use a **stochastic integral** to express the increment of a stochastic process:

$$W(t) - W(s) = \int_s^t dW(u).$$

In words: the increment in Brownian Motion W between s and t is given by summing the instantaneous changes, i.e. the increments over infinitesimal time steps, of Brownian Motion between s and t.

We use Brownian Motion as a building block for the model specification of stock prices, interest rates etc.

For example, we could assume geometric Brownian Motion for a stock price S:

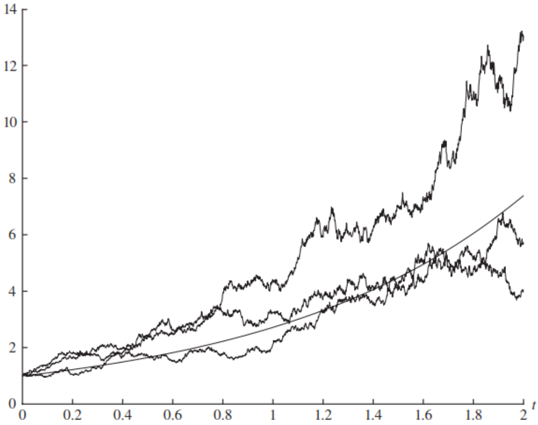
$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t).$$

Or in integral notation:

$$S(t) - S(0) = \int_0^t \mu S(u)du + \int_0^t \sigma S(u)dW(u).$$

This process is called **Geometric Brownian Motion**.

Geometric Brownian Motion with high volatility:



We would like to know the distributional properties of $S(t)$. What can we do?

We could look at the process for $\log S$: apply Itô's Lemma.

Let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x)$, $f_x(t,x)$, and $f_{xx}(t,x)$ are defined and continuous, and let $W(t)$ be a Brownian Motion. Then for every $T > 0$,

$$f(T, W(T)) = f(0, W(0)) + \int_0^T f_t(t, W(t))dt + \int_0^T f_x(t, W(t))dW(t) + \boldsymbol{\frac{1}{2}\int_0^T f_{xx}(t, W(t))dt}.$$

Or (in differential form):

$$df(t) = f_t(t, W(t))dt + f_x(t, W(t))dW(t) + \boldsymbol{\frac{1}{2} f_{xx}(t, W(t))dt}.$$

In contrast to ordinary calculus a second onder term becomes relevant when we deal with Brownian Motion:

$$\frac{1}{2} f_{xx}(t, W(t))(dW(t))^2.$$

The reason is that Brownian Motion paths have non-zero quadratic variation:

$$dW^2(t) = (W(t+dt) - W(t))^2 = dt.$$

The impact of this second order term is deterministic because quadratic variation of a Brownian Motion path over the time interval $[t, t+dt]$ equals dt, regardless of the (random) Brownian path.

Illustration for $f(S(t, W(t))) = \log S$:

$$\begin{aligned} d\log S(t) &= f_t(t, S(t))dt + f_s(t, S(t))dS(t) + \boldsymbol{\frac{1}{2} f_{ss}(t, S(t)) dS^2(t)} \\ &= 0 + \frac{1}{S(t)} dS(t) - \frac{1}{2}\frac{1}{S(t)^2} dS^2(t) \\ &= \frac{1}{S(t)}(\mu S(t)dt + \sigma S(t)dW(t)) - \frac{1}{2}\frac{1}{S(t)^2}(\sigma^2 S^2(t)dt) \\ &= \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW(t). \end{aligned}$$

Or:

$$\log S(t) - \log S(0) = \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma[W(t) - W(0)].$$

Slide deck week 5

Mark-Jan Boes

September, 2020

Part I

Clip 1

In the first week of the course we have looked at pricing an option in the binomial tree model.

Binomial tree model:

- Discrete time
- Finite sample space

But how would it work in a continuous time model where the underlying value of the option can take infinitely many values?

Let us assume that a stock price S follows geometric Brownian Motion:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t),$$

where W is a Brownian Motion under the real world probability measure $\mathbb{P}$.

In the binomial tree model we could solve for the replicating strategy for the option.

The essential part was that, starting at the back of the tree, we could solve a system of linear equations at each node.

We still want to employ the replication method:

- This should be possible because we have just one source of risk
- In addition to the stock we have another asset (the bond or money market account) which follows the process $dB(t) = rB(t)dt$
- With two assets and one source of risk we should be able to replicate all payoffs perfectly (we say we have a complete market)

But the solution method of the binomial tree model does not work because we have infinely many possible payoffs of the option at its maturity.

The idea is that the replicating portfolio of stock and bond will follow **the instantaneous value change of the option exactly**.

Consider an agent who at each time t has a portfolio valued at $X(t)$. This portfolio invests in the money market account and the stock (negative weights are permitted).

Suppose at each time t, the investor holds $\Delta(t)$ shares of stock: $\Delta(t)$ may be random but must be adapted to the filtration associated with the Brownian Motion $W(t), t \geq 0$.

The remainder of the portfolio $X(t) - \Delta(t)S(t)$ is invested in the money market account.

Purpose:

- X(0) equals the option price at time 0, $C(0, S(0))$
- Equate the infinitesimal increments in X and C: $dX = dC$

We can derive the instantaneous change of the portfolio as follows:

$$
\begin{aligned}
dX(t) &= \Delta(t)dS(t) + r(X(t) - \Delta(t)S(t))dt \\
&= rX(t)dt + \Delta(t)(\mu - r)S(t)dt + \Delta(t)\sigma S(t)dW(t).
\end{aligned}
$$

It is important to note that the portfolio is **self-financing**:

- the portfolio value changes due to changes in the market (in this case: changes in the stock price)
- the new portfolio value is reallocated between stocks and bonds
- the portfolio value before and after rebalancing is exactly the same: no external money is necessary to build the replicating portfolio

Now that we have the value development of (replicating) portfolio value X, we also need the value development of the derivative contract C, e.g. a call option written on the underlying value S.

We can derive the differential of $C(t, S(t))$ as follows using Ito's formula:

$$\begin{aligned} dC(t, S(t)) &= \frac{\partial C}{\partial t} dt + \frac{\partial C}{\partial S} dS + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} dS dS \\ &= \left[\frac{\partial C}{\partial t} + \mu S(t) \frac{\partial C}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} \right] dt + \sigma S(t) \frac{\partial C}{\partial S} dW(t). \end{aligned}$$

So, the idea is that the replicating portfolio X starts with some initial capital $X(0)$ and invests in the stock and money market account so that the portfolio value $X(t)$ at each time $t \in [0, T]$ agrees with $C(t, S(t))$

As mentioned before the way to achieve is to:

- choose $X(0) = C(0, S(0))$
- equate the differential equations for X and C

If we first look at the random movements then we see that $\Delta(t)$ should be equal to $\frac{\partial C}{\partial S}$.

Then we need to equate the drift terms:

$$rX(t) + \Delta(t)(\mu - r)S(t) = \frac{\partial C}{\partial t} + \mu S(t)\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 C}{\partial S^2},$$

which is equivalent to:

$$rX(t) - \frac{\partial C}{\partial S} rS(t) = \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 C}{\partial S^2}.$$

Given the initial condition $X(0) = C(0, S(0))$ we can write:

$$rC(t) - \frac{\partial C}{\partial S} rS(t) = \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2(t) \frac{\partial^2 C}{\partial S^2}.$$

Rearranging yields the famous Black-Scholes partial differential equation.

$$\frac{\partial C}{\partial t} + rS(t)\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2(t) \frac{\partial^2 C}{\partial S^2} = rC(t) \quad \text{for all } t \in [0, T).$$

The no-arbitrage call-option price is the solution of this PDE that satisties the terminal condition:

$$C(T, S(T)) = (S(T) - K)^+.$$

Suppose we have found this solution. If an investor starts with initial capital $X(0) = C(0, S(0))$ and uses the hedge $\Delta(t) = \frac{\partial C}{\partial S}$ then the value of the hedge portfolio is exactly equal to the option value for all $t \in [0, T]$.

Hence, we have successfully employed the replication method!

Critically important in this success is that there is just one source of risk driving the randomness in our market and that we have two primary assets (money market and stock): under these assumptions we can replicate all possible payoffs with the two primary assets, in fact derivative contracts are redundant assets in this economy.

Technically, we say that the market is complete (with respect to the money market account and stock).

In their famous 1973 paper, Fisher Black and Myron Scholes showed the solution to the partial differential equation.

For a European call-option (on a non-dividend paying stock):

$$
\begin{aligned}
C(t, S(t)) &= S(t)N(d_1) - Ke^{-r(T-t)}N(d_2) \\
d_1 &= \frac{\ln S - \ln K + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \\
d_2 &= d_1 - \sigma\sqrt{T-t}.
\end{aligned}
$$

With some elementary algebra, you can verify that this solution satisfies the Black-Scholes PDE.

Another way to look at this: suppose a trader has sold the call-option then the instantaneous change in her value is

$$dV = -dC = -\left[\frac{\partial C}{\partial t} + \mu S(t)\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2}\right] dt - \sigma S(t)\frac{\partial C}{\partial S} dW(t).$$

Can she make her portfolio riskless?

Yes, by investing in $\frac{\partial C}{\partial S}$ units of stock:

$$\begin{aligned} d\tilde{V} &= -dC + \frac{\partial C}{\partial S} dS = -dC + \frac{\partial C}{\partial S}(\mu S(t)dt + \sigma S(t)dW(t)) \\ &= -\frac{\partial C}{\partial t} dt - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} dt. \end{aligned}$$

This is a riskless portfolio, so this portfolio should earn the risk free rate (by no arbitrage):

$$d\tilde{V} = r\tilde{V}dt = r\left(-C + \frac{\partial C}{\partial S}S\right)dt.$$

Equating this to the expression for $d\tilde{V}$ of the previous slide yields:

$$r\left(-C + \frac{\partial C}{\partial S}S\right) = -\frac{\partial C}{\partial t}dt - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2}dt,$$

which is again the Black-Scholes PDE.

Part II

Clip 2

In the previous knowledge clip we introduced the **Black-Scholes market**.

This market consists of a stock

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t),$$

where W is a Brownian Motion under the real world probability measure $\mathbb{P}$.

We also have a bond (sometimes referred to as cash or money market account):

$$dB(t) = rB(t)dt.$$

We also assume **frictionless markets**:

- Ability to go long and short for all assets
- Absence of bid-ask spread for all assets
- Absence of transaction costs
- Infinite divisibility of stock and bond positions, i.e. fractional holdings are allowed
- The market is completely liquid, i.e. it is always possible to buy and/or sell unlimited quantities on the market

We have found that within this model framework it is possible to replicate the payoff of an option perfectly by means of a dynamic strategy in stocks and bonds.

In particular, we derived that the option pricing formula, in this model, should satisfy the following partial differential equation:

$$rC(t) - \frac{\partial C}{\partial S} rS(t) = \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 C}{\partial S^2}.$$

The famous Black-Scholes option pricing formula is the solution of this PDE.

Next question: how does the **risk-neutral valuation method** work in this model?

Recall that the no-arbitrage price of a European call option with strike price K and time to maturity T can be found as follows:

$$C(t, S(t)) = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}\left(\max(S(T) - K, 0)\right).$$

In order to be able to evaluate the risk-neutral expectation we need to find the process of S under the risk-neutral probability measure.

Consider a measurable space $(X, \mathcal{F})$ on which there are defined two separate measures μ and ν.

- If, for all $A \in \mathcal{F}$, it holds that

$$\mu(A) = 0 \Rightarrow \nu(A) = 0,$$

  then ν is said to be absolutely continuous with respect to μ on $\mathcal{F}$ and we write this as

$$\nu << \mu.$$

- If we have both $\mu << \nu$ and $\nu << \mu$, then μ and ν are said to be equivalent and we write

$$\mu \sim \nu.$$

Let us look at few examples:

- Consider exponential and log-normal distributions. They both have finite probability for positive values and zero probability for negative values. Hence, these measures are **equivalent**.
- Log-normal distribution is absolutely continuous wrt normal distribution, but not otherwise because normally distributed random variables can take negative values.

There is a **Radon-Nikodym theorem** that relates two measures.

Suppose we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let Z be an almost surely nonnegative random variable with $\mathbb{E}(Z) = 1$. For $A \in \mathcal{F}$, define

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega).$$

Then $\tilde{\mathbb{P}}$ is a probability measure. Furthermore, if X is a nonnegative random variable, then

$$\tilde{\mathbb{E}}(X) = \mathbb{E}(XZ).$$

We can show that if Z is strictly positive, $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are equivalent, i.e. both measures agree on which sets in $\mathcal{F}$ have probability zero.

In financial models:

- We can see the sample space Ω as the set of unknown possible scenarios of the future
- These scenarios have an actual probability measure $\mathbb{P}$
- For the purpose of pricing derivative securities we will use a risk-neutral measure $\tilde{\mathbb{P}}$
- These two measures are equivalent: they **must agree** on what is possible and impossible
- They **may disagree** on how probable the possibilities are
- It is common that we used terminology like *real world* and *risk-neutral world*
- This is a bit unfortunate because there is only one world in the models, represented by the single sample space Ω

Change of measure means that we want to change the distribution of random variable X without changing the possible outcomes of X.

Suppose now that X is a standard normal random variable, i.e.:

$$\mu_X(B) = \mathbb{P}\{X \in B\} = \int_B \phi(x)dx, \quad \forall B \text{ of } \mathbb{R},$$

where

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

is the standard normal density.

If we take $B = (-\infty, b]$ this reduces to something more familiar:

$$\mathbb{P}\{X \leq B\} = \int_{-\infty}^{b} \phi(x)dx, \quad \forall b \in \mathbb{R}.$$

Let θ be a constant and define $Y = X + \theta$, so that under $\mathbb{P}$:

- $\mathbb{E}(Y) = \theta$
- $\mathrm{Var}(Y) = 1$

For the remainder of this example we assume θ to be positive.

We want to change to a new probability measure $\tilde{\mathbb{P}}$ on Ω under which Y is a standard normal random variable.

We don't want to do this by subtracting θ from Y, but by

- assigning less probability to those ω for which $Y(\omega)$ is sufficiently positive
- assigning more probability to those ω for which $Y(\omega)$ is sufficiently negative

Again, we want to change the distribution of Y without changing the possible outcomes of Y.

All the previous has shown us that we need a random variable Z with particular properties to define $\tilde{\mathbb{P}}$.

Let's try:

$$Z(\omega) = \exp\left\{-\theta X(\omega) - \frac{1}{2}\theta^2\right\}$$

It is obvious that $Z(\omega)$ is strictly positive for all $\omega \in \Omega$.

Now we derive $\mathbb{E}(Z)$:

$$\begin{aligned}
\mathbb{E}(Z) &= \int_{-\infty}^{\infty} \exp\left\{-\theta x - \frac{1}{2}\theta^2\right\} \phi(x) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(x^2 + 2\theta x + \theta^2)\right\} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(x + \theta)^2\right\} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}y^2\right\} dy = 1.
\end{aligned}$$

So, random variable Z, which we call the **Radon Nikodym derivative**, has all the necessary characteristics for creating the new probability measure $\tilde{\mathbb{P}}$ through:

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega) \quad \text{for all } A \in \mathcal{F}.$$

The random variable Z has the property that if $X(\omega)$ is positive then $Z(\omega) < 1$.

This shows that $\tilde{\mathbb{P}}$ assigns less probability to than $\mathbb{P}$ to sets on which X is positive, a step in the right direction of statistically recentering Y.

Now we want to look at the probability distribution of Y under $\tilde{\mathbb{P}}$

$$
\begin{aligned}
\tilde{\mathbb{P}}\{Y \leq B\} &= \int_{\{\omega; Y(\omega) \leq b\}} Z(\omega) d\mathbb{P}(\omega) \\
&= \int_{\Omega} \mathbb{I}_{\{Y(\omega) \leq b\}} Z(\omega) d\mathbb{P}(\omega) \\
&= \int_{\Omega} \mathbb{I}_{\{X(\omega) \leq b-\theta\}} \exp\left\{-\theta X(\omega) - \frac{1}{2}\theta^2\right\} d\mathbb{P}(\omega) \\
&= \int_{-\infty}^{\infty} \mathbb{I}_{\{X(\omega) \leq b-\theta\}} \exp\left\{-\theta X(\omega) - \frac{1}{2}\theta^2\right\} \phi(x) dx
\end{aligned}
$$

$$\tilde{\mathbb{P}}\{Y \leq B\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b-\theta} e^{-\theta x - \frac{1}{2}\theta^2} e^{-\frac{x^2}{2}} dx$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b-\theta} e^{-\frac{1}{2}(x+\theta)^2} dx$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-\frac{1}{2}y^2} dy$$

And, hence, Y is standard normal with respect to probability measure $\tilde{\mathbb{P}}$.

Part III

Clip 3

In the previous knowledge clip we talked about a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a nonnegative random variable Z satisfying $\mathbb{E}(Z) = 1$. We then defined a new probability measure $\tilde{\mathbb{P}}$ by the formula:

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega) \quad \forall A \in \mathcal{F}.$$

Or:

$$d\tilde{\mathbb{P}}(\omega) = Z(\omega) d\mathbb{P}(\omega).$$

We say that Z is the **Radon-Nikodym derivative** of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$ and we write:

$$Z(\omega) = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}.$$

In a **finite probability model**, we actually have:

$$Z(\omega) = \frac{\tilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)},$$

which implies

$$\tilde{\mathbb{P}}(A) = \sum_{\omega \in A} Z(\omega)\mathbb{P}(\omega).$$

This formula shows that if we multiply the original probability with an outcome-dependent factor (denoted by Z) we'll get the adjusted probability.

Part IV

Clip 4

In the previous knowledge clip we talked about a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a nonnegative random variable Z satisfying $\mathbb{E}(Z) = 1$. We then defined a new probability measure $\tilde{\mathbb{P}}$ by the formula:

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega) \quad \forall A \in \mathcal{F}.$$

Any random variable X defined on $(\Omega, \mathcal{F}, \mathbb{P})$ now has two expectations:

- one under the probability measure $\mathbb{P}$, which we denote $\mathbb{E}(X)$
- one under the probability measure $\tilde{\mathbb{P}}$, which we denote $\tilde{\mathbb{E}}(X)$

These are related by:

$$\tilde{\mathbb{E}}(X) = \mathbb{E}(XZ).$$

We say that Z is the Radon-Nikodym derivative of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$ and we write:

$$Z(\omega) = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}.$$

In a finite probability model, we actually have:

$$Z(\omega) = \frac{\tilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}.$$

In a general probability model though, we cannot write this formula because $\mathbb{P}(\omega)$ is typically zero for each individual ω.

Now we want to perform a change of measure **in order to change the mean of a whole process** rather than for a single random variable.

Suppose we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathcal{F}(t)$, defined for $0 \leq t \leq T$, where T is a fixed final time. We can define the **Radon-Nikodym derivative process** as:

$$Z(t) = \mathbb{E}[Z|\mathcal{F}(t)], \quad 0 \leq t \leq T$$

We can easily show that the process Z is a martingale:

$$\mathbb{E}[Z(t)|\mathcal{F}(s)] = \mathbb{E}\left[\mathbb{E}[Z|\mathcal{F}(t)]|\mathcal{F}(s)\right] = \mathbb{E}[Z|\mathcal{F}(s)] = Z(s).$$

We can also show for given t satisfying $0 \leq t \leq T$ and Y an $\mathcal{F}(t)$-measurable random variable:

$$\tilde{\mathbb{E}}[Y] = \mathbb{E}[YZ(t)].$$

For $0 \leq s \leq t \leq T$ and Y an $\mathcal{F}(t)$-measurable random variable, it also holds that:

$$\tilde{\mathbb{E}}[Y|\mathcal{F}(s)] = \frac{1}{Z(s)}\mathbb{E}[YZ(t)|\mathcal{F}(s)].$$

Hence, we can calculate expectations and conditional expectations of an $\mathcal{F}(t)$-measurable random variable under a new probability measure $\tilde{\mathbb{P}}$ which is defined by the Radon-Nikodym derivative process.

But it is still not clear how all of this is going to help us in an easy way: **we need something more**.

In general, we can derive **processes** under the new probability measure $\tilde{\mathbb{P}}$ by applying **Girsanov's theorem:**

Let $W(t)$, $0 \leq t \leq T$, be Brownian Motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{F}(t)$, $0 \leq t \leq T$, be a filtration for this Brownian Motion. Let $\Theta(t)$, $0 \leq t \leq T$, be an adapted process and define

$$Z(t) = \exp\left\{-\int_0^t \Theta(u)dW(u) - \frac{1}{2}\int_0^t \Theta(u)^2 du\right\}$$

$$\tilde{W}(t) = W(t) + \int_0^t \Theta(u)du.$$

Set $Z = Z(T)$. Then $\mathbb{E}(Z) = 1$ and **under the new probability measure $\tilde{\mathbb{P}}$, the process $\tilde{W}(t)$ is a Brownian Motion**.

In this theorem:

- It is obvious that $\tilde{W}(t)$ is not a Brownian Motion under $\mathbb{P}$
- If we want to prove that $\tilde{W}(t)$ is a Brownian Motion under $\tilde{\mathbb{P}}$ we should show that it has continuous paths, has quadratic variation t at each time t, and is a martingale under $\tilde{\mathbb{P}}$
- The first two requirements are easily checked
- We can prove that the proces $Z(t), 0 \leq t \leq T$ is a Radon-Nikodym derivative process
- With the results on earlier slides with respect to Z we can prove that $\tilde{W}(t)$ is a martingale under $\tilde{\mathbb{P}}$

By means of using Z we can adjust the original probability measure in such a way that $\tilde{W}(t)$ gets all properties of a Brownian Motion.

Illustration for Geometric Brownian Motion as data generating process:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t),$$

where W is a Brownian Motion under the probability measure $\mathbb{P}$.

In addition we have the money market account (we assume constant interest rates):

$$dB(t) = rB(t)dt.$$

Now, use Ito's lemma to derive the process for the discounted stock price:

$$\begin{aligned} d\left(\frac{S(t)}{B(t)}\right) &= (\mu - r)\frac{S(t)}{B(t)}dt + \sigma\frac{S(t)}{B(t)}dW(t) \\ &= \sigma\frac{S(t)}{B(t)}\left(dW(t) + \frac{\mu - r}{\sigma}dt\right). \end{aligned}$$

Ok, now we can apply Girsanov's theorem to this SDE with:

$$\Theta(t) = \Theta = \frac{\mu - r}{\sigma}.$$

The theorem tells us that:

$$\tilde{W}(t) = W(t) + \left(\frac{\mu - r}{\sigma}\right) t,$$

is a Brownian Motion under $\tilde{\mathbb{P}}$.

So, under $\tilde{\mathbb{P}}$ the discounted stock price process is:

$$d\left(\frac{S(t)}{B(t)}\right) = \sigma \frac{S(t)}{B(t)} d\tilde{W}(t).$$

We can see from the SDE that under $\tilde{\mathbb{P}}$ the **discounted stock price process is a martingale**.

We can again use Ito's lemma to obtain the stock price process under $\tilde{\mathbb{P}}$:

$$dS(t) = rS(t)dt + \sigma S(t)d\tilde{W}(t),$$

where $\tilde{W}$ is a Brownian Motion under the new probability measure $\tilde{\mathbb{P}}$.

Hence, under the $\tilde{\mathbb{P}}$ the stock prices earns the risk-free rate r instead of μ, which was the growth rate under the original probability measure $\mathbb{P}$.

Part V

Clip 5

In the previous knowledge clip we saw that Girsanov's theorem is a tool that can be used to **change the mean of a process**.

If W is a $\mathbb{P}$-Brownian Motion and Θ a process adapted to the filtration $\mathcal{F}$, then

$$\tilde{W}(t) = W(t) + \int_0^t \Theta(u)du,$$

is a $\tilde{\mathbb{P}}$-Brownian Motion.

The adjusted probability measure $\tilde{\mathbb{P}}$ can be found through the Radon-Nikodym derivative process:

$$Z(t) = \exp\left\{-\int_0^t \Theta(u)dW(u) - \frac{1}{2}\int_0^t \Theta(u)^2 du\right\}, \quad t \geq 0.$$

We have applied a trick (Girsanov's theorem) to move from the process under the one probability measure to a process under a new probability measure.

However, we haven't established a **formal connection** with arbitrage pricing yet.

This is where **the first fundamental theorem of asset pricing** comes in.

In the first week of the course of the week we claimed, and illustrated in a discrete setting that **no-arbitrage is equivalent** to

- The existence of a positive **linear pricing rule**
- The existence of a positive **pricing kernel (aka stochastic discount factor)**
- The existence of positive **risk-neutral probabilities**

The heart of the matter is that no arbitrage is equivalent to the existence of a positive pricing rule (partly proven in exercise set 1), the rest follows.

A linear pricing rule maps the possible payoffs of a financial instrument in different futures states of the world to no arbitrage prices:

$$\nu = \eta G,$$

where η is the linear pricing vector, ν the vector of no-arbitrage prices and G a payoff matrix.

The existence of a positive linear pricing rule η ensures that there are no arbitrage opportunities in this model.

In this model we defined **risk-neutral probabilities** as:

$$q_i = \frac{\eta_i}{\sum_{i=1}^m \eta_i}.$$

And for pricing of payoff z we established,

$$\begin{aligned} V(z) &= \sum_{i=1}^m \eta_i z_i = \left(\sum_{i=1}^m \eta_i\right) \sum_{i=1}^m \frac{\eta_i}{\left(\sum_{i=1}^m \eta_i\right)} z_i = \left(\sum_{i=1}^m \eta_i\right) \sum_{i=1}^m q_i z_i \\ &= \left(\sum_{i=1}^m \eta_i\right) \mathbb{E}^{\mathbb{Q}}(z) = \frac{1}{1+r} \mathbb{E}^{\mathbb{Q}}(z). \end{aligned}$$

Hence, the existence of a positive linear pricing rule leads to the existence of adjusted probabilities, which we call risk-neutral probabilities because all assets earn the risk free rate, in expectation, when we use these probabilities.

Well, now in continuous time...

First Fundamental Theorem of Asset Pricing
A financial market is arbitrage-free under the probability measure $\mathbb{P}$ if and only if there exists another probability measure $\mathbb{Q}$, that is equivalent to $\mathbb{P}$, under which all discount asset prices are martingales.

This means that the existence of $\mathbb{Q}$ implies that the time t no arbitrage price of a financial instrument V can be calculated as follows:

$$\frac{V(t)}{B(t)} = \mathbb{E}_t^{\mathbb{Q}}\left(\frac{V(T)}{B(T)}\right),$$

where we use B as the asset with which we discount.

If the process of B is specified as follows:

$$dB(t) = rB(t)dt,$$

then the valuation formula simplifies to:

$$V(t) = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}\left(V(T)\right),$$

Under this assumption for B, we could also say that arbitrage opportunities are ruled out if we can find a probability measure $\mathbb{Q}$, equivalent to $\mathbb{P}$, under which all assets earn the risk free rate, in expectation.

The probability measure $\mathbb{Q}$ is then called **the risk-neutral probability measure**.

In the previous knowledge clip, we derived the process for the discounted stock price in the Black-Scholes world (under $\mathbb{P}$):

$$d\left(\frac{S(t)}{B(t)}\right) = (\mu - r)\frac{S(t)}{B(t)}dt + \sigma\frac{S(t)}{B(t)}dW(t)$$
$$= \sigma\frac{S(t)}{B(t)}\left(dW(t) + \frac{\mu - r}{\sigma}dt\right).$$

Using Girsanov's theorem, we got the discounted stock price process under an adjusted probability measure $\tilde{\mathbb{P}}$:

$$d\left(\frac{S(t)}{B(t)}\right) = \sigma\frac{S(t)}{B(t)}d\tilde{W}(t),$$

$\tilde{W}(t)$ is a Brownian Motion under the new probability measure $\tilde{\mathbb{P}}$.

Using the properties of Brownian Motion, we can conclude that the discounted stock price process is a martingale under $\tilde{\mathbb{P}}$.

Hence, we found a probability measure $\tilde{\mathbb{P}}$ under which the discounted stock price is a martingale: the FFTAP says that (1) the market of stock and bond is free or arbitrage en (2) we can use this probability measure to find no-arbitrage prices of other assets in this market e.g. options.

A European call option with strike price K and time to maturity $(T-t)$, for instance, could be priced in the following way:

$$\frac{C(S(t),t)}{B(t)} = \mathbb{E}_t^{\tilde{\mathbb{P}}}\left(\frac{\max(S(T)-K,0)}{B(T)}\right).$$

Using Ito's lemma we can find that under $\tilde{\mathbb{P}}$ the process of S is given by:

$$dS(t) = rS(t)dt + \sigma S(t)d\tilde{W}(t),$$

where $\tilde{W}$ is a Brownian Motion under the new probability measure $\tilde{\mathbb{P}}$.

Hence, under the $\tilde{\mathbb{P}}$ the stock prices earns the risk-free rate r, in expectation. Therefore we say that $\tilde{\mathbb{P}}$ is the risk-neutral probability measure in this model, i.e. the measure under which all assets earn the risk free rate in expectation.

To summarize:

- Girsanov's theorem is a tool that we can use to make discounted asset prices martingales by changing the original probability measure
- The first fundamental theorem ensures that if we use the new probability measure to price other assets in the market, the resulting prices are free of arbitrage

Part VI

Clip 6

Now that we have knowledge on the First Fundamental Theorem of Asset Pricing, we finally can formally derive the Black-Scholes option pricing formula.

The FFTAP says that under no arbitrage there exists a risk-neutral probability measure, equivalent to the real world probability measure, under which all discounted asset prices are martingales:

$$\frac{V(t)}{B(t)} = \mathbb{E}_t^{\mathbb{Q}}\left(\frac{V(T)}{B(T)}\right).$$

Hence, for a European call option C with strike price K and maturity T we have:

$$C(t, S(t)) = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}} \max(S(T) - K, 0).$$

Hence, the only thing we need to do is to work out the expectation:

$$C(t, S(t)) = e^{-r(T-t)} \int_0^\infty \max(s(T) - K, 0) q(s(T)) ds(T).$$

We know that in the Black-Scholes the stock price follows a Geometric Brownian Motion under the risk-neutral probability measure:

$$\begin{aligned} S(T) &= e^{\log S(T)} \\ &= \exp\left\{\log S(t) + \left(r - \frac{1}{2}\sigma^2\right)(T-t) + \sigma\left(W(T) - W(t)\right)\right\} \\ &= S(t)\exp\left\{\left(r - \frac{1}{2}\sigma^2\right)(T-t) + \sigma\left(W(T) - W(t)\right)\right\} \end{aligned}$$

We rewrite the integral in terms of Y, a random variable with a normal distribution.

$$C(t, S(t)) = e^{-r(T-t)} \int_{-\infty}^{\infty} \max(S(t)e^{y(T)} - K, 0) q(y(T)) dy(T).$$

We get rid of the max function by adjusting the interval over which we are integrating:

$$C(t, S(t)) = e^{-r(T-t)} \int_{\log \frac{K}{S(t)}}^{\infty} (S(t)e^{y(T)} - K) q(y(T)) dy(T).$$

We can split the integral in two parts:

$$C(t, S(t)) = e^{-r(T-t)} \int_{\log \frac{K}{S(t)}}^{\infty} S(t) e^{y(T)} q(y(T)) dy(T)$$
$$- K e^{-r(T-t)} \int_{\log \frac{K}{S(t)}}^{\infty} q(y(T)) dy(T).$$

Let us first focus on the second integral. We can transform to a standard normal random variable Z:

$$Ke^{-r(T-t)} \int_{\frac{\log \frac{K}{S(t)} - (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}}^{\infty} q(z)dz.$$

Using the properties of the standard normal distribution, we obtain:

$$Ke^{-r(T-t)} \int_{-\infty}^{\frac{\log \frac{S(t)}{K} + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}} q(z)dz = Ke^{-r(T-t)} N(d_2),$$

where

$$d_2 = \frac{\log \frac{S(t)}{K} + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}.$$

Now, we turn to the first integral:

$$e^{-r(T-t)} \int_{\log \frac{K}{S(t)}}^{\infty} S(t)e^{y(T)} q(y(T))dy(T).$$

Again, we want to make the transformation to the standard normal random variable Z

$$e^{-r(T-t)} \int_{\frac{\log \frac{K}{S(t)} - (r-\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}}^{\infty} S(t)e^{z\sigma\sqrt{(T-t)}+(r-\frac{1}{2}\sigma^2)(T-t)} q(z)dz.$$

This equals:

$$e^{-r(T-t)}e^{(r-\frac{1}{2}\sigma^2)(T-t)}S(t)\int_{-d_2}^{\infty} e^{z\sigma\sqrt{(T-t)}}q(z)dz.$$

Then:

$$e^{-r(T-t)}e^{(r-\frac{1}{2}\sigma^2)(T-t)}S(t)\int_{-d_2}^{\infty} e^{z\sigma\sqrt{(T-t)}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz.$$

Using the properties of the normal distribution we get:

$$e^{-r(T-t)}e^{(r-\frac{1}{2}\sigma^2)(T-t)}S(t)\int_{-\infty}^{d_2} e^{-z\sigma\sqrt{(T-t)}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz.$$

This can be rewritten as:

$$e^{-r(T-t)}e^{\frac{1}{2}\sigma^2(T-t)}e^{(r-\frac{1}{2}\sigma^2)(T-t)}S(t)\int_{-\infty}^{d_2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(z+\sigma\sqrt{(T-t)}\right)^2}dz.$$

Another substitution gives:

$$e^{-r(T-t)}e^{\frac{1}{2}\sigma^2(T-t)}e^{(r-\frac{1}{2}\sigma^2)(T-t)}S(t)\int_{-\infty}^{d_2+\sigma\sqrt{(T-t)}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}h^2}dh,$$

which equals:

$$S(t)\int_{-\infty}^{d_1}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}h^2}dh = S(t)N(d_1),$$

where

$$d_1 = \frac{\log\frac{S(t)}{K} + (r+\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}.$$

So, taking it all together we arrive at the Black-Scholes formula for a European call option:

$$C(t, S(t)) = S(t)N(d_1) - Ke^{-r(T-t)}N(d_2).$$

Part VII

Summary

This course is about finding the **no-arbitrage price** of **derivatives** contracts.

No-arbitrage pricing means that prices are determined in such a way that there is **no free lunch** in the market.

Björk defines an **arbitrage portfolio** as a portfolio h with the properties:

$$V_0^h = 0$$
$$V_1^h > 0 \text{ with probability } 1.$$

An arbitrage portfolio is thus a deterministic money making machine.

This week we mainly worked in **the Black-Scholes world**.

The Black-Scholes market consists of a **stock**

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t),$$

where W is a Brownian Motion under the real world probability measure $\mathbb{P}$.

We also have a **bond** (sometimes referred to as cash or money market account):

$$dB(t) = rB(t)dt.$$

We also assume **frictionless markets**:

- Ability to go long and short for all assets
- Absence of bid-ask spread for all assets
- Absence of transaction costs
- Infinite divisibility of stock and bond positions, i.e. fractional holdings are allowed
- The market is completely liquid, i.e. it is always possible to buy and/or sell unlimited quantities on the market

Under these assumptions, we derived the time t no-arbitrage price of a European call option with strike price K and time to maturity $(T - t)$, using two different methods:

- the replication method
- the risk-neutral valuation method

In the **replication method**, we tried to setup a portfolio of stock and bond such that the instantaneous change of the replicating portfolio value matches the instantaneous change of the call option, i.e. $dX = dC$.

We found the following SDE for the **replication portfolio process**:

$$\begin{aligned} dX(t) &= \Delta(t)dS(t) + r(X(t) - \Delta(t)S(t))dt \\ &= rX(t)dt + \Delta(t)(\mu - r)S(t)dt + \Delta(t)\sigma S(t)dW(t). \end{aligned}$$

And the following SDE for the **call option process**:

$$\begin{aligned} dC(t, S(t)) &= \frac{\partial C}{\partial t}dt + \frac{\partial C}{\partial S}dS(t) + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}dS(t)dS(t) \\ &= \left[\frac{\partial C}{\partial t} + \mu S(t)\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S(t)^2 \frac{\partial^2 C}{\partial S^2}\right] dt + \sigma S(t)\frac{\partial C}{\partial S}dW(t). \end{aligned}$$

If we choose $\Delta(t) = \frac{\partial C}{\partial S}$ then the impact of the instantaneous increment in Brownian Motion is the same on X and C.

Next step is to equate the drift terms of both processes (we use initial condition $X(0) = C(0, S(0))$):

$$rC(t) - \frac{\partial C}{\partial S} rS(t) = \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 C}{\partial S^2}.$$

Rearranging yields the famous Black-Scholes partial differential equation.

$$\frac{\partial C}{\partial t} + rS(t)\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 C}{\partial S^2} = rC(t) \quad \text{for all } t \in [0, T).$$

The no-arbitrage price should be the solution of this partial differential equation: with the solution of this equation, evaluated at time t, as cash in our hand, we are able to construct a dynamic strategy of stock and bond that replicates the option value perfectly.

In their famous 1973 paper, Fisher Black and Myron Scholes showed the solution to the partial differential equation.

For a European call-option (on a non-dividend paying stock):

$$\begin{aligned} C(t, S(t)) &= S(t)N(d_1) - Ke^{-r(T-t)}N(d_2) \\ d_1 &= \frac{\ln S - \ln K + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \\ d_2 &= d_1 - \sigma\sqrt{T-t}. \end{aligned}$$

With some elementary algebra, you can verify that this solution satisfies the Black-Scholes PDE.

A second method to determine the no-arbitrage of a European call option is the **risk-neutral valuation method**:

$$C(t, S(t)) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left(C(T, S(T))|\mathcal{F}_t\right),$$

where $\mathbb{Q}$ is a probability measure under which all assets earn the risk free rate.

In the **first week of the course**, we have shown that we are allowed to do this (in a discrete setting):

- in the setting of a binomial tree
- by referring to fundamental theorems of finance: no arbitrage is equivalent to the existence of a positive linear pricing kernel is equivalent to the existence of a risk-neutral probability measure.

We use **Girsanov's theorem** to make the change of probability measure, i.e. to change the mean, for a stochastic process:

If W is a $\mathbb{P}$-Brownian Motion and Θ a process adapted to the filtration $\mathcal{F}(t)$, then

$$\tilde{W}(t) = W(t) + \int_0^t \Theta(u)du,$$

is a $\tilde{\mathbb{P}}$-Brownian Motion.

The adjusted probability measure $\tilde{\mathbb{P}}$ can be found through the **Radon-Nikodym derivative process**:

$$Z(t) = \exp\left\{-\int_0^t \Theta(u)dW(u) - \frac{1}{2}\int_0^t \Theta(u)^2 du\right\}, \quad t \geq 0.$$

In the Black-Scholes world:

$$d\left(\frac{S(t)}{B(t)}\right) = (\mu - r)\frac{S(t)}{B(t)}dt + \sigma\frac{S(t)}{B(t)}dW(t)$$
$$= \sigma\frac{S(t)}{B(t)}\left(dW(t) + \frac{\mu - r}{\sigma}dt\right).$$

Using Girsanov's theorem, we can get the discounted stock price process under an adjusted probability measure $\tilde{\mathbb{P}}$:

$$d\left(\frac{S(t)}{B(t)}\right) = \sigma\frac{S(t)}{B(t)}d\tilde{W}(t),$$

$\tilde{W}(t)$ is a Brownian Motion under the new probability measure $\tilde{\mathbb{P}}$.

Using the properties of Brownian Motion, we can conclude that the discounted stock price process is a martingale under $\tilde{\mathbb{P}}$.

Hence, we found a probability measure $\tilde{\mathbb{P}}$ under which the discounted stock price is a martingale: the **First Fundamental Theorem of Asset Pricing** now says that (1) the market of stock and bond is free or arbitrage en (2) we can use this probability measure to find no-arbitrage prices of other assets in this market e.g. options.

A European call option with strike price K and time to maturity $(T-t)$, for instance, could therefore be priced in the following way:

$$\frac{C(S(t),t)}{B(t)} = \mathbb{E}_t^{\tilde{\mathbb{P}}}\left(\frac{\max(S(T)-K,0)}{B(T)}\right).$$

Using Ito's lemma we can find that under $\tilde{\mathbb{P}}$ the process of S is given by:

$$dS(t) = rS(t)dt + \sigma S(t)d\tilde{W}(t),$$

where $\tilde{W}$ is a Brownian Motion under the new probability measure $\tilde{\mathbb{P}}$.

Hence, under the $\tilde{\mathbb{P}}$ the stock prices earns the risk-free rate r, in expectation. Therefore we say that $\tilde{\mathbb{P}}$ is **the risk-neutral probability measure** in this model.

Slide deck week 6

Mark-Jan Boes

October, 2020

Part I

Knowledge clip 1

Last week, we derived the **no-arbitrage price** of a European call option in the Black-Scholes world by applying **Girsanov's Theorem** and **the First Fundamental Theorem of Asset Pricing**:

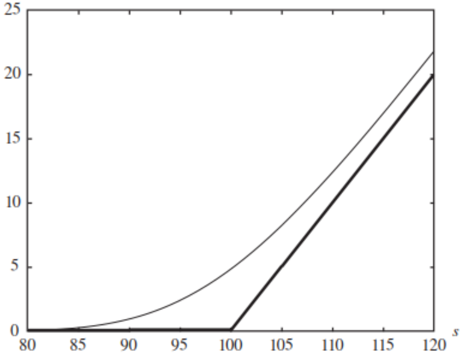
$$C(t, S(t)) = S(t)N(d_1) - Ke^{-r(T-t)}N(d_2),$$

with:

$$d_1 = \frac{\log\left(\frac{S(t)}{K}\right) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = d_1 - \sigma\sqrt{T-t}.$$

Graphical illustration of call-option price with strike $K = 100$:



- Let $V(t, S)$ be the value of a portfolio consisting of stock S and options on it.
- It is practically important to know not only the value of the portfolio, but also its sensitivity to
  - Price changes of the underlying
  - Changes of model parameters
- In other words, it is not only important to know how much the portfolio worth, but also how much **risk** do you have in your portfolio.
- In the first case above, we are interested in how much our portfolio changes when the market moves.
- In the second case, we are interested in how much our portfolio value changes if we recalibrate our model, which in fact is again sensitivity to the market but indirect.

The sensitivities have standard notation with Greek letters (that is why call the sensitivities "The Greeks").

$$\Delta = \frac{\partial V}{\partial S} \quad \text{Delta}$$

$$\Gamma = \frac{\partial^2 V}{\partial S^2} \quad \text{Gamma}$$

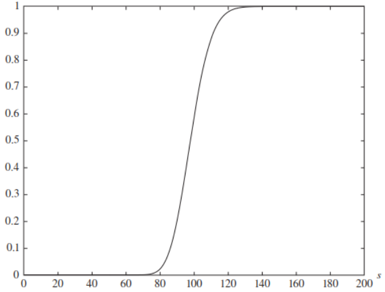
$$\rho = \frac{\partial V}{\partial r} \quad \text{Rho}$$

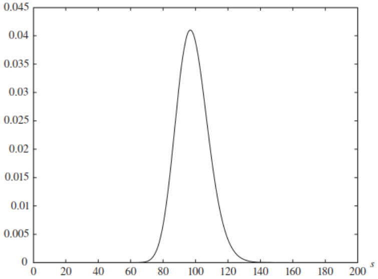
$$\Theta = \frac{\partial V}{\partial t} \quad \text{Theta}$$

$$\mathcal{V} = \frac{\partial V}{\partial \sigma} \quad \text{Vega}$$

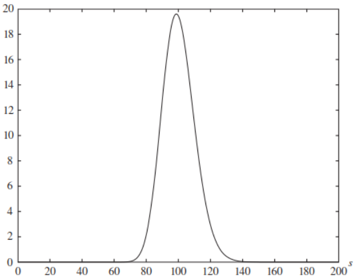
- If the portfolio is insensitive to *small* changes in the market values or parameters, it is called **risk neutral**.
- If it is insensitive to changes in the stock price, it means that it has zero Delta, and it is called **delta neutral**.
- The procedure to reduce or remove risks in the portfolio is called **hedging**.
- But let us first look at the Greeks self, in the Black–Scholes model.

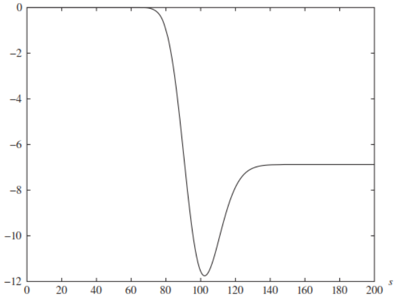
Delta and Gamma





Vega and Theta





In the Black–Scholes model the Greeks of a long position in a European call option are given by the formulas

$$\Delta = \mathcal{N}(d_1)$$

$$\Gamma = \frac{\phi(d_1)}{S\sigma\sqrt{T-t}}$$

$$\rho = K(T-t)e^{-r(T-t)}\mathcal{N}(d_2)$$

$$\Theta = -\frac{S\phi(d_1)\sigma}{2\sqrt{T-t}} - rKe^{-r(T-t)}\mathcal{N}(d_2)$$

$$\mathcal{V} = S\phi(d_1)\sqrt{T-t}$$

Call option delta in a Black-Scholes world.

$$C_t = S_t N(d_1) - K\exp(-r(T-t))N(d_2)$$

$$\frac{\partial C}{\partial S} = N(d_1) + S_t \frac{\partial N(d_1)}{\partial S} - K\exp(-r(T-t))\frac{\partial N(d_2)}{\partial S}$$

$$\frac{\partial C}{\partial S} = N(d_1) + S_t \frac{\partial N(d_1)}{\partial d_1}\frac{\partial d_1}{\partial S} - K\exp(-r(T-t))\frac{\partial N(d_2)}{\partial d_2}\frac{\partial d_2}{\partial S}$$

$$\frac{\partial N(d_1)}{\partial d_1} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} d_1^2\right)$$

$$\begin{aligned}
\frac{\partial N(d_2)}{\partial d_2} &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(d_1 - \sigma\sqrt{T-t})^2\right) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} d_1^2\right) \exp(d_1 \sigma \sqrt{T-t} - \frac{1}{2}\sigma^2(T-t)) \\
&= \frac{\partial N(d_1)}{\partial d_1} \exp(\ln(S_t/K) + r(T-t)) \\
&= \frac{\partial N(d_1)}{\partial d_1} \frac{S_t \exp(r(T-t))}{K}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial C}{\partial S} &= N(d_1) + S_t \frac{\partial N(d_1)}{\partial d_1} \frac{1}{\sigma\sqrt{T-t}} \cdots \\
\cdots & - \frac{\partial N(d_1)}{\partial d_1} \frac{K\exp(-r(T-t))S_t\exp(r(T-t))}{K} \\
&= N(d_1) + \frac{\partial N(d_1)}{\partial d_1} \frac{1}{S_t\sigma\sqrt{T-t}} - \frac{\partial N(d_1)}{\partial d_1} \frac{1}{S_t\sigma\sqrt{T-t}} \\
&= N(d_1)
\end{aligned}
$$

Hence, delta is given by the cumulative distribution function of the standard normal distribution evaluated in d_1

This means that delta is a number between 0 and 1.

The lower the strike, the more in-the-money the call option is, the higher delta.

The higher the strike, the more out-of-the-money the call option is, the lower delta.

- In practice, traders seldom take an open position, like a naked option.
- They and risk managers prefer a position with low or zero sensitivity to market moves.
- So they **hedge** their position.
- If $V(t,S)$ is the portfolio to hedge and $F(t,S)$ is a hedge instrument, the total new portfolio is

$$\bar{V}(t,S) = V(t,S) + x \cdot F(t,S)$$

and we choose x such that

$$\frac{\partial V}{\partial S}(t,S) + x\frac{\partial F}{\partial S}(t,S) = 0$$

the *hedge ratio* is thus

$$x = -\frac{\Delta_V}{\Delta_F}$$

- You can keep your portfolio delta neutral only for *small* market movements (Gamma typically is not zero!).
- What if market moves substantially?
- Then you would not have your portfolio risk neutral and you will have large value changes if you are long or short Gamma.
- If you do not want such surprises or you expect large movements, you may hedge also your Gamma. This will guarantee that your portfolio stays longer risk neutral.
- You need to add a security with non-zero Gamma, like an option.

$$\bar{V}(t,S) = V(t,S) + x_F \cdot F(t,S) + x_G \cdot G(t,S)$$

and we choose x such that

$$\frac{\partial \bar{V}}{\partial S}(t,S) = 0 \quad \text{and} \quad \frac{\partial^2 \bar{V}}{\partial S^2}(t,S) = 0$$

to find x_F and x_G.

Part II

Knowledge clip 2

In the Black-Scholes market we did the following:

- We proposed a model under $\mathbb{P}$
- With the FFTAP in the back of our mind we derived the process for the discounted stock price process
- We concluded that the drift of the resulting process, $\Theta(t)$, satisfies the conditions of Girsanov's theorem
- We applied Girsanov's theorem to get the process for the discounted stock price under $\tilde{\mathbb{P}}$
- Concluded that this was a martingale and hence, $\tilde{\mathbb{P}}$ is a risk-neutral measure
- Applied the FFTAP to ensure that the model didn't allow for arbitrage opportunities
- Using the properties of the risk-neutral measure and the FFTAP that the no-arbitrage price of a payoff $X(T)$ is given by:

$$X(t) = e^{-r(T-t)}\mathbb{E}_t^{\tilde{\mathbb{P}}}(X(T)).$$

We took the FFTAP **as a given** and then it was easy to claim that the no-arbitrage price of a payoff $X(T)$ is given by:

$$X(t) = e^{-r(T-t)}\mathbb{E}_t^{\tilde{\mathbb{P}}}(X(T)).$$

Some text books (including Shreve) take a slightly different path to establish this result. Let's go investigate this different path that builds on **the replication method**, **Girsanov's theorem** and **the Martingale Representation Theorem**.

Hence, the approach does not make use of the FFTAP that we (conveniently) assumed to be true.

Shreve defines the discount process, using the adapted interest rate process $R(t)$ as:

$$D(t) = e^{-\int_0^t R(s)ds}.$$

The discounted stock price process is given by:

$$d(D(t)S(t)) = \sigma D(t)S(t)[\Theta(t)dt + dW(t)],$$

where we define the **market price of risk** to be:

$$\Theta(t) = \frac{\mu - R(t)}{\sigma}.$$

This reflects the compensation for each unit of risk in this particular model: each source of risk (in this model we have only one) might lead to a required compensation of risk, depending on risk preferences.

Consider an agent who begins with capital $X(0)$ and at each time t, $0 \leq t \leq T$, holds $\Delta(t)$ shares of stock, investing or borrowing at the interest rate $R(t)$ as necessary to finance this.

For Geometric Brownian Motion the differential of this agent's portfolio is given by:

$$\begin{aligned} dX(t) &= \Delta(t)dS(t) + R(t)(X(t) - \Delta(t)S(t))dt \\ &= \Delta(t)(\mu S(t)dt + \sigma S(t)dW(t)) + R(t)(X(t) - \Delta(t)S(t))dt \\ &= R(t)X(t)dt + \Delta(t)\sigma(t)S(t)\left[\left(\frac{\mu - R(t)}{\sigma}\right)dt + dW(t)\right]. \end{aligned}$$

With Itô's product rule we can derive that:

$$d(D(t)X(t)) = \Delta(t)\sigma(t)D(t)S(t)\left[\left(\frac{\mu - R(t)}{\sigma}\right)dt + dW(t)\right]$$
$$= \Delta(t)d(D(t)S(t)).$$

Changes in the discounted value of an agent's portfolio are entirely due to fluctuations in the discounted stock price. Using Girsanov's theorem, we arrive at:

$$d(D(t)X(t)) = \Delta(t)\sigma(t)D(t)S(t)d\tilde{W}(t).$$

Our agent has two investment options:

- a money market account with mean rate of return r
- a stock with mean rate or return r

Regardless of how the agent invests, the portfolio will earn the risk free rate r (in expectation), ie, the **discounted value** must be a martingale.

Let $V(T)$ be an $\mathcal{F}(T)$-measurable random variable that represents the payoff at time T of a derivative security.

We want to know what initial capital $X(0)$ and portfolio strategy $\Delta(t), 0 \leq t \leq T$, an agent would need to in order to have,

$$X(T) = V(T) \quad \text{almost surely.}$$

We will see in a second that this can be done. Once it has been done, we use the fact that $D(t)X(t)$ is a martingale under $\tilde{\mathbb{P}}$:

$$D(t)X(t) = \tilde{\mathbb{E}}[D(T)X(T)|\mathcal{F}(t)] = \tilde{\mathbb{E}}[D(T)V(T)|\mathcal{F}(t)].$$

The value $X(t)$ is the capital needed at time t in order to successfully complete the replication. Hence, we can call this the price $V(t)$ of the derivative security at time t:

$$D(t)V(t) = \tilde{\mathbb{E}}[D(T)V(T)|\mathcal{F}(t)].$$

In the most general formulation this becomes:

$$V(t) = \tilde{\mathbb{E}}\left[e^{-\int_t^T R(u)du} V(T)|\mathcal{F}(t)\right] \quad 0 \leq t \leq T.$$

This result critically depends on the assumption that if an agent begins with the correct initial capital, there is a portfolio process $\Delta(t), 0 \leq t \leq T$ such that the agent's portfolio value at the final time T will be $V(T)$ almost surely.

This is where the **Martingale Representation Theorem** comes in.

Martingale Representation Theorem Let $W(t), 0 \leq t \leq T$ be a Brownian Motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{F}(t)$ be the filtration generated by this Brownian Motion. Let $M(t)$ be a martingale with respect to this filtration. Then there is an adapted process $\Gamma(u), 0 \leq u \leq T$ such that:

$$M(t) = M(0) + \int_0^t \Gamma(u) dW(u), \quad 0 \leq t \leq T$$

Or in differential notiation:

$$dM(t) = \Gamma(t) dW(t).$$

How to apply this theorem? We already have established that discounted asset prices (also derivatives) are martingales under $\tilde{\mathbb{P}}$:

$$D(t)V(t) = \tilde{\mathbb{E}}[D(T)V(T)|\mathcal{F}(t)].$$

Hence, the theorem says:

$$D(t)V(t) = V(0) + \int_0^t \tilde{\Gamma}(u)d\tilde{W}(u), \quad 0 \leq t \leq T.$$

On the other hand, for any portfolio process $\Delta(t)$, the discounted portfolio value is given by:

$$D(t)X(t) = X(0) + \int_0^t \Delta(u)\sigma(u)D(u)S(u)d\tilde{W}(u).$$

In order to have $X(t) = V(t)$ for all t, we should choose

$$X(0) = V(0),$$

and choose $\Delta(t)$ to satisfy:

$$\Delta(t)\sigma(t)D(t)S(t) = \tilde{\Gamma}(t),$$

which is equivalent to:

$$\Delta(t) = \frac{\tilde{\Gamma}(t)}{\sigma(t)D(t)S(t)}.$$

The Martingale Representation Theorem justifies the risk-neutral pricing formulas but it does not provide a practical method of finding the replicating portfolio $\Delta(t)$.

Hence, the Martingale Representation Theorem guarantees that a process $\tilde{\Gamma}$ exists and therefore a replicating strategy exists, but it does not provide a method for finding $\tilde{\Gamma}(t)$.

Part III

Knowledge clip 3

Last week, we introduced the **First Fundamental Theorem of Asset Pricing**:

A financial market is **arbitrage-free** under the probability measure $\mathbb{P}$ if and only if there exists another probability measure $\mathbb{Q}$, that is equivalent to $\mathbb{P}$, under which all **discounted asset prices are martingales**.

The probability measure $\mathbb{Q}$ is called the **equivalent martingale measure**.

If this equivalent martingale measure has the property that, under this probability measure, all tradeable assets earn the risk free rate (in expectation), then this measure is called the **risk-neutral probability measure**.

Given the crucial importance of this theorem for everything we do in quant finance, I try to outline the proof in these slides.

Let us recall what an arbitrage opportunity is: An **arbitrage** possibility on a financial market is a **self-financing** portfolio h such that

$$V^h(0) = 0$$
$$\mathbb{P}(V^h(T) \geq 0) = 1$$
$$\mathbb{P}(V^h(T) > 0) > 0.$$

In words: an arbitrage possibility is a way of trading so that one starts with 0 capital and at some later time T is sure not to have lost money **and** has a positive probability of having made money.

What we want to do: assume the existence of a risk-neutral probability measure and then show that arbitrage possibilities cannot exist.

By definition we know that the existence of a risk-neutral measure means that all discounted asset prices follow a martingale process under this risk-neutral measure.

This should **also** hold for a **portfolio value**, as a portfolio value results directly from the asset prices.

Shreve introduces the discount process in his book (we use constant interest rates over here, in line with the Black-Scholes market):

$$D(t) = e^{-\int_0^t r du} = e^{-rt}.$$

Given the properties of the risk-neutral measure we know that the portfolio value process satisfies (martingale property):

$$\tilde{\mathbb{E}}(D(T)V(T)) = D(0)V(0) = V(0).$$

Now the link to the definition of arbitrage: Let $V(t)$ be a portfolio process with $V(0) = 0$. Then we have

$$\tilde{\mathbb{E}}(D(T)V(T)) = 0.$$

Suppose that the portfolio V leads to an arbitrage then by definition of arbitrage we know:

$$\mathbb{P}(V(T) < 0) = 0.$$

I.e. we can be sure that in the real world the portfolio does not generate any losses.

Equivalence between $\mathbb{P}$ and $\tilde{\mathbb{P}}$ implies:

$$\tilde{\mathbb{P}}(V(T) < 0) = 0.$$

But we already know that the portfolio is not delivering any money in excess of the risk-free rate (in expectation) under $\tilde{\mathbb{P}}$ (see previous slide):

$$\tilde{\mathbb{E}}(D(T)V(T)) = 0.$$

The previous two formulas imply:

$$\tilde{\mathbb{P}}(V(T) > 0) = 0.$$

Because of equivalence between $\mathbb{P}$ and $\tilde{\mathbb{P}}$:

$$\mathbb{P}(V(T) > 0) = 0.$$

Hence, no arbitrage.

So, all the previous allows us to price derivative contracts using the following formula:

$$X(t) = e^{-r(T-t)}\mathbb{E}_t^{\tilde{\mathbb{P}}}(X(T)).$$

Using the relationship between $\mathbb{P}$ and $\tilde{\mathbb{P}}$ (through the Radon-Nikodym derivative Z), we can be sure now that this price can also be calculated as:

$$X(t) = e^{-r(T-t)}\mathbb{E}_t^{\tilde{\mathbb{P}}}(X(T)) = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{P}}\left(X_T\frac{Z(T)}{Z(t)}\right).$$

If we then define the pricing kernel π as:

$$\pi(t) = e^{-rt} Z(t) = \exp\left\{-rt - \int_0^t \Theta(u) dW(u) - \frac{1}{2}\int_0^t \Theta(u)^2 du\right\},$$

Then,

$$X(t) = \mathbb{E}_t^{\mathbb{P}}\left(X(T)\frac{\pi(T)}{\pi(t)}\right).$$

Given that Z is positive, π is also positive meaning that a positive pricing kernel implies no-arbitrage: a result that we already established in week 1 of the course in a discrete setting.

We discussed the First Fundamental Theorem of Asset Pricing: the existence of a probability measure equivalent to the real world probability measure, under which all discounted asset prices are martingales implies no-arbitrage.

Second Fundamental Theorem of Asset Pricing: Consider a market model that has a risk-neural probability measure, i.e. does not allow for arbitrage opportunities. The model is complete if and only if the risk-neutral probability measure is unique.

We need to know what completeness means: a market model is complete if every derivative security can be hedged, i.e a model in which options are redundant assets.

When are options redundant assets? Well, if those options can be replicated perfectly: basically, we wouldn't need the options then.

The Black-Scholes world is such a world and therefore the Second Fundamental Theorem implies that the risk-neutral probability measure is unique.

Suppose we would model the stock price by means of the following stochastic differential equation:

$$dS(t) = \mu S(t)dt + \sigma_1 S(t)dW_1(t) + \sigma_2 S(t)dW_2(t).$$

From the equity literature we know that perhaps even 7 factors drive equity returns, so an extension of the Black-Scholes world like the above SDE seems to be a good next step?

Is this market complete?

The answer is no:

- The call option price is driven by two Brownian Motions, two sources of risk
- The value of the replicating strategy of cash and stocks is also driven by these two sources of randomness
- However, I can just choose one position in the stock (this is $\Delta(t)$ in our Black-Scholes analysis), so the replicating strategy follows the movements of W_1 or W_2 or a bit of both
- Hence, the replicating strategy won't be able to track the option price perfectly

What is the implication of the incompleteness of this model? We won't be able to apply the Law of One Price.

There are more prices for the option that do not allow for arbitrage, i.e. the no-arbitrage price of an option is not uniquely defined.

Can we make this market complete? Yes, by adding a second stock that depends on at least one of the two Brownian Motions.

For completeness:

- count the number of sources of randomness
- the number of assets in the replicating strategy should be this number $+ 1$

Part IV

Knowledge clip 4

Suppose we have the price of a call-option available, for instance by applying the Black-Scholes formula.

Question: what can we say about the fair price of the put-option?

The payoff of the European call-option at maturity is given by:

$$\Phi(S(T)) = (S(T) - K)^+.$$

And the payoff function of the European put-option is:

$$\Phi(S(T)) = (K - S(T))^+.$$

Suppose now that we would take a long position in a call and a short position in a put:

$$\Phi(S(T)) = (S(T) - K)^+ - (K - S(T))^+ = S(T) - K.$$

Hence, the typical option payoff disappears. Notice that the resulting payoff is the payoff of a forward contract with reference price K.

The values of the left hand side should then also be equal to the right hand side at all previous times:

$$C(t, S(t)) - P(t, S(t)) = e^{-r(T-t)}\mathbb{E}_t^Q(S(T) - K),$$

which equals:

$$C(t, S(t)) - P(t, S(t)) = S(t) - Ke^{-r(T-t)}$$

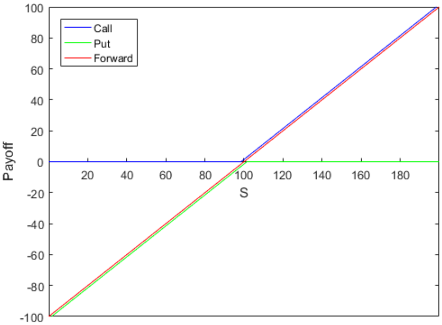
Rewriting this yields:

$$P(t, S(t)) = C(t, S(t)) - S(t) + Ke^{-r(T-t)}.$$

Having knowledge on the call-option price immediately gives the fair value of the put-option.

Note that we have established this relationship without using any models, i.e. this relationship is model-independent.

Graphically:



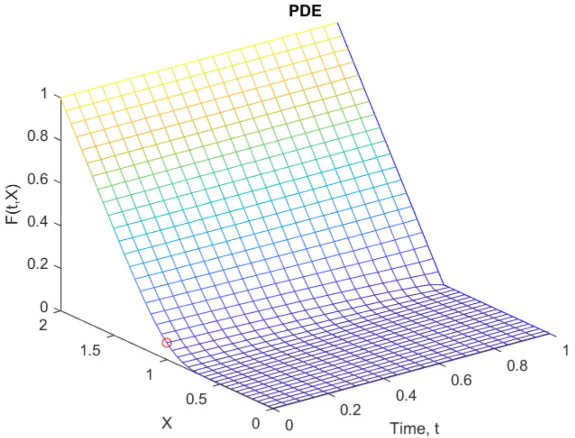
Remember that we started our Black-Scholes analysis with the derivation of the Black-Scholes PDE:

$$rC(t) - \frac{\partial C}{\partial S} rS(t) = \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 C}{\partial S^2},$$

that satisfies the terminal condition:

$$C(T, S(T)) = (S(T) - K)^+.$$

Can we find a solution for this equation? Here is where the Feynman-Kac theorem comes in.


PDE
F(t,X)
0
0.2
0.4
0.6
0.8
1
2
1.5
1
0.5
0
X
0
0.2
0.4
0.6
0.8
1
Time, t

The Feynman–Kac theorem says that the solution of the Partial Differential Equation (PDE)

$$\frac{\partial F}{\partial t}(t,x) + \mu(t,x)\frac{\partial F}{\partial x}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 F}{\partial x^2}(t,x) = 0$$

with the boundary condition

$$F(T,x) = \Phi(x).$$

is given by the expectation

$$F(t,X_t) = \mathbb{E}_t\left[\Phi(X_T)\right]$$

where X satisfies the SDE

$$\begin{aligned} dX_s &= \mu(s,X_s)ds + \sigma(s,X_s)dW_s \\ X(t) &= x. \end{aligned}$$

Consider the following PDE

$$\frac{\partial F}{\partial t}(t,x) + \frac{1}{2}\sigma^2 \frac{\partial^2 F}{\partial x^2}(t,x) = 0$$
$$F(T,x) = x^2$$

From Feynman-Kac theorem we have that

$$F(t,x) = E_t[X_T^2]$$

where

$$dX_t = 0 \cdot dt + \sigma dW_t, \quad X_t = x$$

We know the solution

$$X_T = x + \sigma(W_T - W_t)$$

Then

$$F(t,x) = E_t[X_T^2] = Var_t[X_T] + (E_t[X_T])^2 = \sigma^2(T-t) + x^2$$

And the discounted version of the theorem:

$$\frac{\partial F}{\partial t}(t,x) + \mu(t,x)\frac{\partial F}{\partial x}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 F}{\partial x^2}(t,x) - \mathrm{r}\,\mathrm{F}(\mathrm{t},\mathrm{x}) = 0$$
$$F(T,x) = \Phi(x)$$

which has the solution

$$F(t,X_t) = \mathbb{E}_t\left[\mathrm{e}^{-\mathrm{r}(\mathrm{T}-\mathrm{t})}\Phi(X_T)\right]$$

with the corresponding SDE

$$dX_s = \mu(s,X_s)ds + \sigma(s,X_s)dW_s$$
$$X(t) = x$$

We had the Black-Scholes PDE:

$$rC(t) - \frac{\partial C}{\partial S} rS(t) = \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2(t) \frac{\partial^2 C}{\partial S^2},$$

with terminal condition:

$$C(T, S(T)) = (S(T) - K)^+.$$

Rewriting the PDE in Feynman-Kac-form:

$$\frac{\partial C}{\partial t} + \frac{\partial C}{\partial S} rS(t) + \frac{1}{2}\sigma^2 S^2(t) \frac{\partial^2 C}{\partial S^2} - rC(t) = 0.$$

The theorem says that the call option price is given by:

$$C(t, S(t)) = \mathbb{E}_t\left[e^{-r(T-t)}(S(T)-K)^+)\right],$$

where the expectation is calculated using the properties of the following SDE:.

$$dS(t) = rS(t)dt + \sigma S(t)dW(t),$$

where W is a Brownian Motion process.

Yes, indeed, this becomes the very same expectation that we worked out in earlier knowledge clips.

Now, we didn't use the FFTAP but we used the Black-Scholes PDE combined with the Feynman-Kac theorem.

Feynman-Kac theorem:

- The theorem shows us how partial differential equations can be solved with stochastic differential equations
- But also the other way around: we can solve SDEs by using PDEs
- For almost every problem, we have two methods to (numerically) attack the problem:
  - We can use **Monte Carlo simulation** on the basis of the provided **stochastic differential equation**
  - We can write the corresponding **partial differential equation** and solve it, using for example finite difference methods

Part V

Knowledge clip 5

In the Black-Scholes world, we have ...

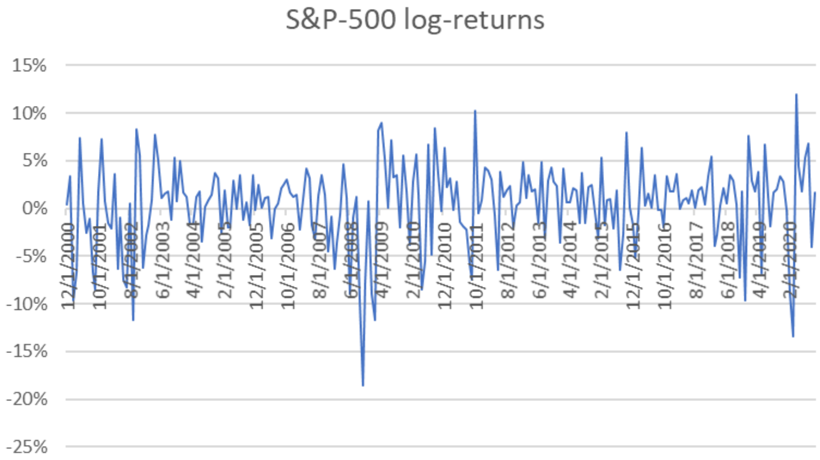
- A market
  - Bank account $dB(t) = rB(t)dt$
  - Stock price $dS(t) = \mu S(t)dt + \sigma S(t)dW(t)$
  - The process W is a Brownian Motion under the probability measure $\mathbb{P}$.
  - The market is frictionless
- A risk premium for taking risk in W equal to $\mu - r$
- Market price of risk: $\frac{\mu-r}{\sigma}$
- Distribution of log-returns (under $\mathbb{P}$):

$$\log S(t) - \log S(0) \sim N\left((\mu - \frac{1}{2}\sigma^2)t, \sigma^2 t\right).$$

Question:
Is the Black-Scholes model a model that describes empirical data properly?

There are a number of ways to check this:

- Use time series data to assess whether log-returns on stock (or stock indices) are normally distributed.
- Use empirical option data and check whether each price fits with the same volatility of the underlying

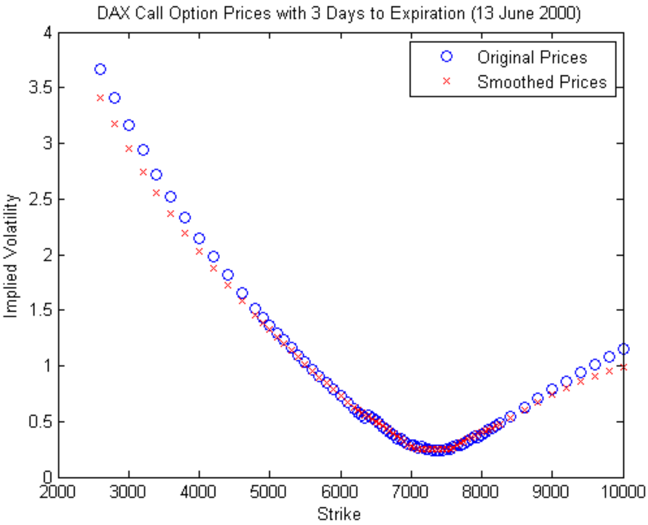

S&P-500 log-returns
15%
10%
5%
0%
-5%
-10%
-15%
-20%
-25%
12/1/2000
10/1/2001
8/1/2002
6/1/2003
4/1/2004
2/1/2005
12/1/2005
10/1/2006
8/1/2007
6/1/2008
4/1/2009
2/1/2010
12/1/2010
10/1/2011
8/1/2012
6/1/2013
4/1/2014
2/1/2015
12/1/2015
10/1/2016
8/1/2017
6/1/2018
4/1/2019
2/1/2020

The data:

- 20 years of monthly log-returns on the S&P-500 index
- Skewness is -0.83
- Kurtosis is 4.75
- Jarque-Bera test: null hypothesis of normality
- p-value of this test is 0.00: null hypothesis is rejected at very small significance levels

Conclusion: empirical data do not fit with the assumptions of the Black-Scholes market.

Empirical data



The data:

- We took market data of option prices
- We inverted the Black-Scholes market to calculate the value of volatility that fits with option data
- We call these: implied volatilities (implied by the Black-Scholes formula)
- Did this for several strike prices (maturity data fixed)

Conclusion: empirical data do not fit with the assumptions of the Black-Scholes market.

Ultimate purpose:
Find a model that fits with all properties of empirical data: returns data as well as option data (all strikes, all maturities).

First step to take from the Black-Scholes model: **introduce stochastic volatility**.

Heston stochastic volatility model:

$$
\begin{aligned}
dS(t) &= \mu S(t)dt + \sigma(t)S(t)dW^S(t) \\
d\sigma^2(t) &= -\kappa(\sigma^2(t) - \sigma^2)dt + \sigma_\sigma \sigma(t)dW^V(t) \\
d[W^S, W^V]_t &= \rho dt.
\end{aligned}
$$

The last equation implies that the correlation between instantaneous changes in W^S and W^V have correlation equal to ρ.

Note that variance is stochastic but is not a tradeable asset: we have a model with two tradeable assets (stock and bond) and two sources of risk.

How could we possibly do option pricing in this model?

I really hope that now, at the end of this course, you have an idea:

- Replication method
- Risk-neutral valuation method
- Pricing kernel method

Let us try **risk-neutral valuation**.

In order to apply this method, we first rewrite the model in terms of **independent Brownian Motions**:

$$dS(t) = \mu S(t)dt + \sigma(t)S(t)dW^S(t)$$
$$d\sigma^2(t) = -\kappa(\sigma^2(t) - \sigma^2)dt + \sigma_\sigma \sigma(t)\left(\rho dW^S(t) + \sqrt{1-\rho^2}dW^V(t)\right).$$

To check this:

- Evaluate $Cov_t(dS(t), d\sigma^2(t))$ in both models
- Evaluate $Var_t(d\sigma^2(t))$ in both models

We could rewrite the formulas as follows:

$$dS(t) = rS(t)dt + \sigma(t)S(t)\left\{dW^S(t) + \left(\frac{\mu - r}{\sigma(t)}\right)dt\right\}$$

$$\begin{aligned} d\sigma^2(t) &= -(\kappa + \eta^V)\left(\sigma^2(t) - \frac{\kappa\sigma^2}{\kappa + \eta^V}\right)dt \\ &\quad + \sigma_\sigma\sigma(t)\left(\rho\left\{dW^S(t) + \frac{\mu - r}{\sigma(t)}dt\right\}\right. \\ &\quad \left. + \sqrt{1-\rho^2}\left\{dW^V(t) + \frac{1}{\sqrt{1-\rho^2}}\left(\frac{\eta^V\sigma_t}{\sigma_\sigma} - \rho\left(\frac{\mu - r}{\sigma}\right)\right)dt\right\}\right). \end{aligned}$$

Now use **Girsanov's theorem** to get the model under **the risk-neutral probability measure**:

$$
\begin{aligned}
dS(t) &= rS(t)dt + \sigma(t)S(t)d\tilde{W}^S(t) \\
d\sigma^2(t) &= -(\kappa+\eta^V)\left(\sigma^2(t) - \frac{\kappa\sigma^2}{\kappa+\eta^V}\right)dt \\
&\quad + \sigma_\sigma\sigma(t)\left(\rho d\tilde{W}^S(t) + \sqrt{1-\rho^2}d\tilde{W}^V(t)\right).
\end{aligned}
$$

We know that $\tilde{W}^S$ and $\tilde{W}^V$ are both Brownian Motion processes under the **the risk-neutral probability measure** .

In the same spirit as for the Geometric Brownian Motion process:

- The change of measure leads to a change of the drift in the process
- This holds for both the stock price S and the variance process $\sigma^2(t)$
- The volatility of the stock price has not changed after changing the probability measure
- The same holds for the variance process $\sigma^2(t)$
- Conclusion: probabilistically, a change of measure leads to a change in the expectation of the process

The **variance risk premium** is given by $\eta^V \sigma^2(t)$: the difference in the drift between the original probability measure and the risk-neutral probability measure.

How to calculate the no-arbitrage price of a call-option in this model?

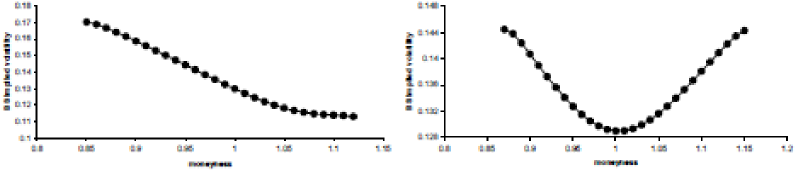
Apply:

$$C(S(t), t) = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}(\max(S(T) - K, 0).$$

In the Heston-model you do that either by:

- Deriving the price analytically (like we did in the knowledge clips for the Black-Scholes model)
- Deriving the price by means of Monte Carlo simulation (like we did in the second computer assignment)

Implied volatility patterns that the Heston-model can generate:



(a) Negative correlation

(b) Zero correlation

Part VI

Summary

This course is about finding the **no-arbitrage price** of **derivatives** contracts.

No-arbitrage pricing means that prices are determined in such a way that there is **no free lunch** in the market.

Björk defines an **arbitrage portfolio** as a portfolio h with the properties:

$$V_0^h = 0$$
$$V_1^h > 0 \text{ with probability } 1.$$

An arbitrage portfolio is thus a deterministic money making machine.

The Black-Scholes market consists of a **stock**

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t),$$

where W is a Brownian Motion under the real world probability measure $\mathbb{P}$.

We also have a **bond** (sometimes referred to as cash or money market account):

$$dB(t) = rB(t)dt.$$

We also assume **frictionless markets**:

- Ability to go long and short for all assets
- Absence of bid-ask spread for all assets
- Absence of transaction costs
- Infinite divisibility of stock and bond positions, i.e. fractional holdings are allowed
- The market is completely liquid, i.e. it is always possible to buy and/or sell unlimited quantities on the market

Last week, we derived the **no-arbitrage price** of a European call option in the Black-Scholes world by applying **Girsanov's Theorem** and **the First Fundamental Theorem of Asset Pricing**:

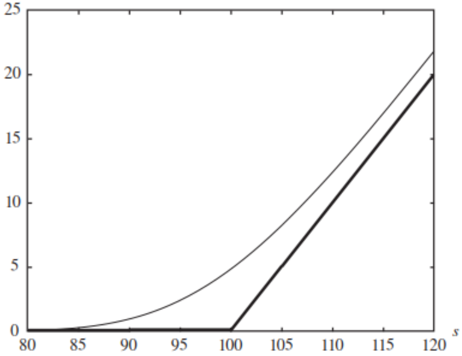
$$C(t, S(t)) = S(t)N(d_1) - Ke^{-r(T-t)}N(d_2),$$

with:

$$d_1 = \frac{\log\left(\frac{S(t)}{K}\right) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = d_1 - \sigma\sqrt{T-t}.$$

Graphical illustration of call-option price with strike $K = 100$:



Suppose now that a **trader** has a portfolio of options e.g. the trader has sold a European call option.

In a practical setting the trader would have sold the option against the **ask price** and tries to make sure that at the option maturity date she replicates the payoff of the option: the difference between the ask-price and the true price (i.e. the mid price) is then guaranteed P&L.

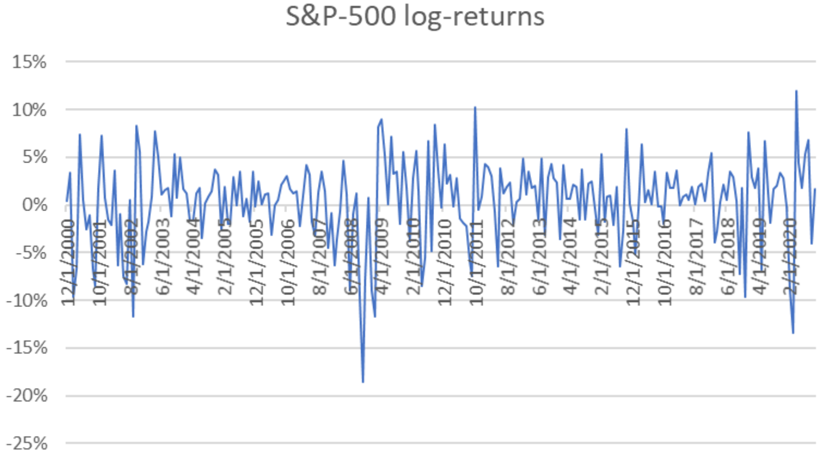
Hence, the trader is not only concerned about the price of the option at trade date but also (very) concerned about how the option value can change between trade date and maturity date.

In the Black–Scholes model the Greeks of a long position in a European call option are given by the formulas

$$
\begin{aligned}
\Delta &= \frac{\partial C}{\partial S} = \mathcal{N}(d_1) \\
\Gamma &= \frac{\partial^2 C}{\partial S^2} = \frac{\phi(d_1)}{S\sigma\sqrt{T-t}} \\
\rho &= \frac{\partial C}{\partial r} = K(T-t)e^{-r(T-t)}\mathcal{N}(d_2) \\
\Theta &= \frac{\partial C}{\partial t} = -\frac{S\phi(d_1)\sigma}{2\sqrt{T-t}} - rKe^{-r(T-t)}\mathcal{N}(d_2) \\
\mathcal{V} &= \frac{\partial C}{\partial \sigma} = S\phi(d_1)\sqrt{T-t}
\end{aligned}
$$

An option trader would keep a close eye on all of these sensitivities: in a Black-Scholes world hedging Δ is sufficient.

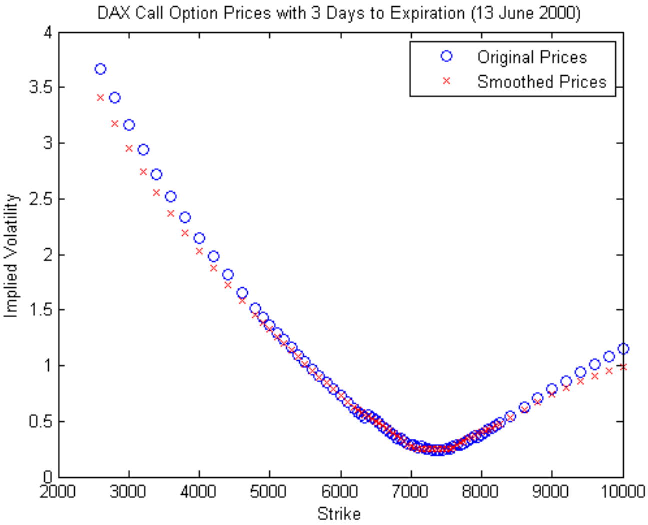
But do we live in a Black-Scholes world? Let's check the data!



The data:

- 20 years of monthly log-returns on the S&P-500 index
- Skewness is -0.83
- Kurtosis is 4.75
- Jarque-Bera test: null hypothesis of normality
- p-value of this test is 0.00: null hypothesis is rejected at very small significance levels

Conclusion: empirical data do not fit with the assumptions of the Black-Scholes market.


DAX Call Option Prices with 3 Days to Expiration (13 June 2000)
Original Prices
Smoothed Prices
Implied Volatility
Strike
4
3.5
3
2.5
2
1.5
1
0.5
0
2000
3000
4000
5000
6000
7000
8000
9000
10000

The data:

- We took market data of option prices
- We inverted the Black-Scholes market to calculate the value of volatility that fits with option data
- We call these: implied volatilities (implied by the Black-Scholes formula)
- Did this for several strike prices (maturity data fixed)

Conclusion: empirical data do not fit with the assumptions of the Black-Scholes market.

Ultimate purpose:
Find a model that fits with all properties of empirical data: returns data as well as option data (all strikes, all maturities).

First step to take from the Black-Scholes model: **introduce stochastic volatility**.

Heston stochastic volatility model:

$$
\begin{aligned}
dS(t) &= \mu S(t)dt + \sigma(t)S(t)dW^S(t) \\
d\sigma^2(t) &= -\kappa(\sigma^2(t) - \sigma^2)dt + \sigma_\sigma \sigma(t)dW^V(t) \\
d[W^S, W^V]_t &= \rho dt.
\end{aligned}
$$

The last equation implies that the correlation between instantaneous changes in W^S and W^V have correlation equal to ρ.

Note that variance is stochastic but is not a tradeable asset: we have a model with two tradeable assets (stock and bond) and two sources of risk.

How could we possibly do option pricing in this model?

I really hope that now, at the end of this course, you have an idea:

- Replication method
- Risk-neutral valuation method
- Pricing kernel method

Let us try **risk-neutral valuation**.

We could use **Girsanov's theorem** to get the model under **the risk-neutral probability measure**:

$$
\begin{aligned}
dS(t) &= rS(t)dt + \sigma(t)S(t)d\tilde{W}^S(t) \\
d\sigma^2(t) &= -(\kappa + \eta^V)\left(\sigma^2(t) - \frac{\kappa\sigma^2}{\kappa + \eta^V}\right)dt \\
&\quad + \sigma_\sigma \sigma(t)\left(\rho d\tilde{W}^S(t) + \sqrt{1-\rho^2}\, d\tilde{W}^V(t)\right).
\end{aligned}
$$

We know that $\tilde{W}^S$ and $\tilde{W}^V$ are both Brownian Motion processes under the **the risk-neutral probability measure** .

How to calculate the no-arbitrage price of a call-option in this model?

Apply:

$$C(S(t),t) = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}(\max(S(T)-K,0).$$

In the Heston-model you do that either by:

- Deriving the price analytically (like we did in the knowledge clips for the Black-Scholes model)
- Deriving the price by means of Monte Carlo simulation (like we did in the second computer assignment)