

In the mathematical finance literature, most articles written in the last two and half decades start with the words "Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, ...". Whether it is explicitly mentioned or not in the articles or texts, the probability space is always there somewhere in the background. The triple  $(\Omega, \mathcal{F}, \mathbb{P})$ , is called a *probability space*.and its inclusion in quantitative finance literature reflects the increasing influence of probability theory and probability theorists over the subject. It forms the foundation of the *probabilistic universe*.

This probability space comprises of three components

1. the sample space  $\Omega$
2. the filtration  $\mathcal{F}$
3. the probability measure  $\mathbb{P}$ .

$$R_i = \begin{cases} +1 & \text{Head} \\ -1 & \text{Tail} \end{cases} <$$

**Example** (3-period binomial model). Take  $n = 3$ ; so  $\Omega = \Omega_3$ , given by the finite set

*Probability space on a coin toss*

$$\Omega = \{UUU, UUD, UDU, UDD, DUU, DUD, DDU, DDD\},$$

Or the set of all possible outcomes of three coin tosses.

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Now define the following two subsets of  $\Omega$  :

$$H: A_U = \{\underline{U}UU, UUD, UDU, UDD\}, \quad T: A_D = \{\underline{D}UU, DUD, DDU, DDD\}.$$

We see  $A_U$  is the subset of outcomes where a Head appears on the first throw,  $A_D$  is the subset of outcomes where a Tail lands on the first throw.

Define also

$$\begin{aligned} A_{\underline{UU}} &= \{\underline{UU}U, UUD\}, & A_{\underline{UD}} &= \{\underline{UD}U, UDD\}, \\ A_{\underline{DU}} &= \{\underline{DU}U, DUD\}, & A_{\underline{DD}} &= \{\underline{DD}U, DDD\} \end{aligned}$$

corresponding to the events that the first two coin tosses result in  $HH$ ,  $HT$ ,  $TH$ ,  $TT$  respectively. At the end of each period, new information becomes available to help us predict the actual stock trajectory.

*→  $\mathcal{F}$  is a mathematical concept, we don't use it in real life*

The filtration is an indication of how information about a probabilistic experiment builds up over time as more results become available. It can be thought of as an increasing family of events. The pair  $(\Omega, \mathcal{F})$  is called a measurable space.

*set of outcomes*

*storage of info and how it builds up*

*• Filtration allows us to record and store the build of information  $(\Omega, \mathcal{F}, \mathbb{R})$*



*previous data is the filtration*

*⇒ will lead onto option pricing as stock price is a martingale*

We place a *probability measure*  $\mathbb{P}$  on  $\{\Omega, \mathcal{F}\}$ .  $\mathbb{P}$  is a special type of "function", called a measure which assigns probabilities to subsets (i.e. the outcomes); the theory also comes from Measure Theory. Whereas cumulative density functions (CDF) are defined on intervals such as  $\mathbb{R}$ , probability measures are defined on general sets, giving greater power, generalisation and flexibility. A probability measure  $\mathbb{P}$  is a function mapping  $\mathbb{P} : \mathcal{F} \longrightarrow [0, 1]$  with the properties

→ CDF

→ (i)  $\mathbb{P}(\Omega) = 1$ ,  
Probability of everything happening  
= 1

↓  
 $\mathbb{P}$  assuming all previous outcomes in the filtration have an outcome between 0, 1

Important to understand what  $\mathbb{P}$ ,  $\mathcal{F}$  and  $\Omega$  denote

(ii) if  $A_1, A_2, \dots$  is a sequence of disjoint sets in  $\mathcal{F}$ , then  $\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) =$

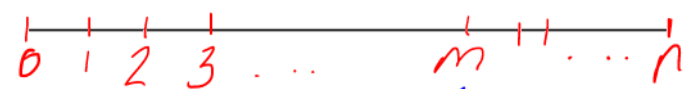
$$\sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

The same as saying probability of x OR y  
∴ add probabilities

↑ Union

# Conditional Expectations

What makes a conditional expectation different (from an unconditional one) is information (just as in the case of conditional probability). In our probability space,  $(\Omega, \mathcal{F}, \mathbb{P})$  information is represented by the filtration  $\mathcal{F}$ ; hence a conditional expectation with respect to the (usual information) filtration seems a natural choice.



$\mathcal{F}_m$   
= past  
information

discrete  $\therefore$  no time  $t$   
future  
 $\mathbb{E}_m[W_n | \mathcal{F}_m]; n > m$   
 $\leftarrow$  step no.

means "stochastic process  $W_n$  such that or conditional upon record of info up to  $m$ "

is the expected value of the random variable conditional upon the filtration set  $\mathcal{F}$ . In general  $W$  will be adapted to the filtration  $\mathcal{F}$ .

We're trying to find future value  $n$  based on  $\mathcal{F}_m$  where  $n$  is stochastic hence  $W$

## Adapted (Measurable) Process

A stochastic process  $S_t$  is said to be adapted to the filtration  $\mathcal{F}_t$  (or measurable with respect to  $\mathcal{F}_t$ , or  $\mathcal{F}_t$ -adapted) if the value of  $S_t$  at time  $t$  is known given the information set  $\mathcal{F}_t$ .



As we collect information and get expectations at intervals we can make predictions about where we may be in the future, past  $t$

Conditional expectations have the following useful properties: If  $X, Y$  are integrable random variables and  $\alpha, \beta$  are constants then we list the following results:

### 1. Linearity:

$$\mathbb{E}[\alpha X + \beta Y | \mathcal{F}] = \alpha \mathbb{E}[X | \mathcal{F}] + \beta \mathbb{E}[Y | \mathcal{F}]$$

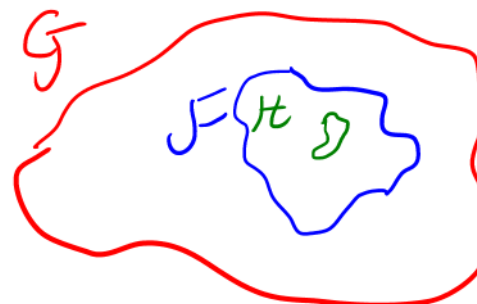
### 2. Tower Property (i.e. Iterated Expectations): if $\mathcal{F} \subset \mathcal{G}$

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

*as we refine our sets, the smaller set wins*

This property states that if taking iterated expectations with respect to several levels of information, we may as well take a single expectation subject to the smallest set of available information. The moral being "the smallest filtration always wins".

*Access to a larger set  $\mathcal{G}$  but as we get more precise, say  $\mathcal{F}$  is a subset of  $\mathcal{G}$  then  $\mathcal{G}$  adds no further information. Then we get to  $\mathcal{F}$  and as the smallest filtration we say it is the one with the most information in it.*



3. As a special case of the Tower property, we have

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}]] = \mathbb{E}[X]$$

since “no filtration” is always a smaller information set than any filtration


*X is  $\mathcal{F}$  adapted*

4. **Taking Out What Is Known:** if  $X$  is  $\mathcal{F}$ -measurable, then the value of  $X$  is known once we know  $\mathcal{F}$ . Therefore,

$$\mathbb{E}[X|\mathcal{F}] = X \quad \because X \text{ is like a scalar and not random}$$

5. **Taking Out What Is Known (2):** by extension, if  $X$  is  $\mathcal{F}$ -measurable but not  $Y$ , *we can take  $X$  out of the conditional expectation*

$$\mathbb{E}[XY|\mathcal{F}] = X\mathbb{E}[Y|\mathcal{F}]$$

*take out* 

6. **Independence:** if  $X$  is independent from  $\mathcal{F}$ , then knowing  $\mathcal{F}$  is useless to predict the value of  $X$ . Hence, *we just have a normal expectation not a conditional one*

$$\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$$

7. **Positivity:** if  $X \geq 0$  then  $\mathbb{E}[X|\mathcal{F}] \geq 0$ .

*used in module 3*

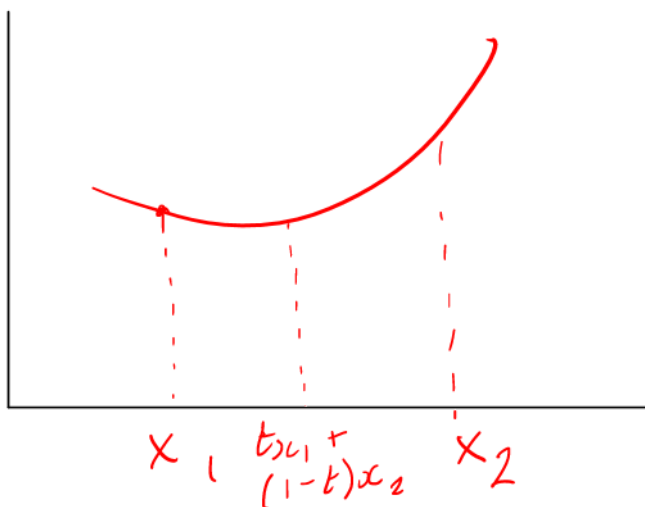
8. **Jensen's Inequality:** First formally define *convexity*. A function  $g : I \rightarrow \mathbb{R}$  is a convex function if for  $x_1, x_2$ , in  $I$  and any  $\lambda \in [0, 1]$

$$g(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda g(x_1) + (1 - \lambda)g(x_2)$$

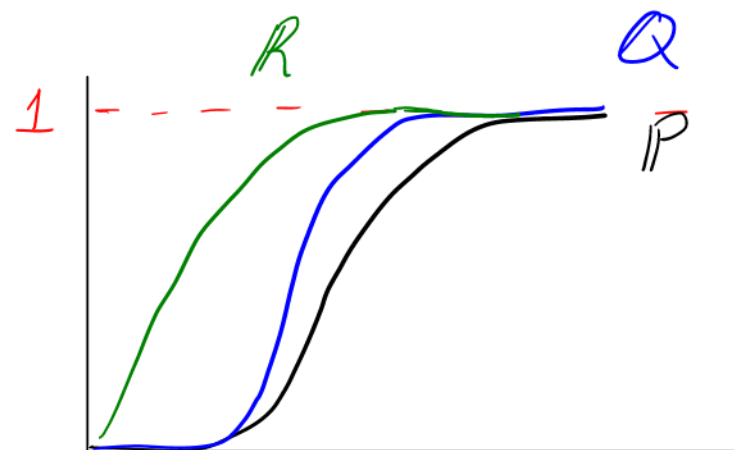
If  $f$  is a convex function, then

$$f(\mathbb{E}[X|\mathcal{F}]) \leq \mathbb{E}[f(X)|\mathcal{F}]$$

*f: convex function*



*What's impossible in  $\mathbb{P}$  (physical world) = What's impossible in  $\mathbb{Q}$  (risk neutral world) but  $\neq \mathbb{P}$ , some other world  $\therefore \mathbb{P} \sim \mathbb{Q}$*





# Changing Probability Measure

You have seen in the Binomial Model lecture that there is more than just one probability measure.

Indeed, the lecture introduced you to the distinction between the “real” or “physical” probability measure, which we encounter every day on our Bloomberg or Reuters screen, and the so-called “risk-neutral” measure, which is used for pricing.

Probability measures are by no means unique. We will see in the next lecture that the powerful arsenal of martingale techniques enables us, under certain assumptions, to change measure and transpose our problem subject to the real world measure into an equivalent problem formulated as a martingale under a different measure.

For now, we just outline the rules that allow us to define equivalent measures.

$\mathbb{P}$  = physical world  
 $\mathbb{Q}$  = risk neutral world

## Equivalent Measure

If two measures  $\mathbb{P}$  and  $\mathbb{Q}$  share the same sample space  $\Omega$  and if  $\mathbb{P}(A) = 0$  implies  $\mathbb{Q}(A) = 0$  for all subset  $A$ , we say that  $\mathbb{Q}$  is **absolutely continuous** with respect to  $\mathbb{P}$  and denote this by  $\mathbb{Q} \ll \mathbb{P}$ .

The key point is that all impossible events under  $\mathbb{P}$  remain impossible under  $\mathbb{Q}$ . The probability mass of the possible events will be distributed differently under  $\mathbb{P}$  and  $\mathbb{Q}$ . In short "it is alright to tinker with the probabilities as long as we do not tinker with the (im)possibilities"

If  $\mathbb{Q} \ll \mathbb{P}$  and  $\mathbb{P} \ll \mathbb{Q}$  then the two measures are said to be **equivalent**, denoted by  $\mathbb{P} \sim \mathbb{Q}$ .

$\therefore$  Impossible in  $\mathbb{P}$  world and impossible in  $\mathbb{Q}$  world means the measures are "equivalent"

Options : start in real world  $\rightarrow$  change to risk neutral world  $\rightarrow$  price option