

# Introduction to Stochastic Processes

Niels O. Nygaard

## 1 Measure and Integration Theory

### 1.1 $\sigma$ -algebras

Consider a set  $\Omega$ . We consider a system of subsets  $\mathcal{F}$  of  $\Omega$ . We assume that  $\mathcal{F}$  satisfies the following axioms

- i)* The set  $\Omega$  itself and the empty set  $\emptyset$  are in  $\mathcal{F}$
- ii)* If the countably (finitely or infinitely) many sets  $\{A_i\}_{i=1}^{\infty}$  are in  $\mathcal{F}$  then the union  $\bigcup A_i$  is in  $\mathcal{F}$
- iii)* If  $A \in \mathcal{F}$  then the complement  $A^c = \Omega - A$  is in  $\mathcal{F}$

Such a system of subsets is called a  $\sigma$ -algebra.

We use the notation  $A \in \mathcal{F}$  to indicate that the set  $A$  is in  $\mathcal{F}$ . Remark that since  $A \cap B = (A^c \cup B^c)^c$  and more generally  $\bigcap A_i = (\bigcup A_i^c)^c$  (*Exercise: Prove this equality*) axiom *ii)* holds with union replaced by intersection. Also we have  $A - B = A \cap B^c$  so if  $A, B \in \mathcal{F}$  then  $A - B \in \mathcal{F}$ .

Basically you should think of a  $\sigma$ -algebra as a system of subsets in which you can perform any of the usual set-theoretic operations (union, intersection, difference) on countably many sets. We get an obvious example by taking  $\mathcal{F}$  to be the system of all subsets of  $\Omega$ ,  $\mathcal{P}(\Omega)$ . At the other extreme the system consisting only of  $\Omega$  itself and  $\emptyset$  is a  $\sigma$ -algebra.

Given any system of subsets  $\mathcal{S}$  of  $\Omega$  we can consider the smallest  $\sigma$ -algebra containing  $\mathcal{S}$ . This is the intersection of all the  $\sigma$ -algebras containing  $\mathcal{S}$ . Since the set of all subsets of  $\Omega$  is a  $\sigma$ -algebra containing  $\mathcal{S}$ , there is at least one  $\sigma$ -algebra containing  $\mathcal{S}$ . The smallest  $\sigma$ -algebra containing  $\mathcal{S}$  is called the  $\sigma$ -algebra generated by  $\mathcal{S}$ .

**Example 1.1** Consider the set of all half-open intervals  $\mathcal{S} = \{[a, b) \mid a, b \in \mathbb{R}, a < b\}$ . Let  $\mathcal{B}$  denote the  $\sigma$ -algebra generated by  $\mathcal{S}$ . The sets in  $\mathcal{B}$  are called the Borel sets. The  $\sigma$ -algebra  $\mathcal{B}$  contains all open and all closed intervals, indeed  $(a, b) = \bigcup_n^\infty [a + \frac{1}{n}, b)$  and  $[a, b] = \bigcap_n^\infty [a, b + \frac{1}{n})$ . Borel sets can be extremely complicated as the following example shows.

**Example 1.2** Consider the closed interval  $[0, 1]$ , this is a Borel set. We remove the open interval  $(\frac{1}{3}, \frac{2}{3})$ . We are left with the union  $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ . From each of these two intervals we remove the middle third  $(\frac{1}{9}, \frac{2}{9})$  and  $(\frac{7}{9}, \frac{8}{9})$ . This leaves us with the union of four intervals  $[0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$ . We continue this way, at each step removing the middle half of each of the intervals and taking the intersection of these countably many Borel sets we end up with a Borel set called the Cantor set. Remark that all the endpoints of the intervals are in the Cantor set, hence it is certainly not empty. The Cantor set is in fact an uncountable set.

## 1.2 Measures

Let  $(\Omega, \mathcal{F})$  be a set and  $\mathcal{F}$  a  $\sigma$ -algebra of subsets of  $\Omega$ . A measure on  $(\Omega, \mathcal{F})$  is a function  $\mu : \mathcal{F} \rightarrow \mathbb{R}_{\geq 0}$  (one sometimes allows  $\mu$  to take the value  $\infty$ ) satisfying the following condition

( $\sigma$ -additivity) If  $A$  is the union of finitely or countably infinitely many pairwise disjoint sets  $A = \bigcup_{i=1}^\infty A_i$  with  $A_i \in \mathcal{F}$  and  $A_i \cap A_j = \emptyset$  when  $i \neq j$  then  $\mu(A) = \sum_{i=1}^\infty \mu(A_i)$ .

Remark that if  $A \subset B$  we have  $\mu(A) \leq \mu(B)$  because  $B = A \cup (B - A)$  so  $\mu(B) = \mu(A) + \mu(B - A)$  and since  $\mu(B - A) \geq 0$  we have  $\mu(B) \geq \mu(A)$

You might think of the measure  $\mu$  as a kind of area or volume or, as we shall be mostly concerned with, as a probability of an event.

**Example 1.3** Consider half-open intervals in  $\mathbb{R}$ . We define  $\mu([a, b)) = b - a$  (the length of the interval). One can extend  $\mu$  to a measure on  $\mathcal{B}$ , the  $\sigma$ -algebra of Borel sets.

Some properties of  $\mu$ :

$\mu([a, b]) = b - a$ , indeed we have  $[a, b] \subset [a, b] \subset [a, b + \frac{1}{n}]$  for all  $n$ . Thus  $b - a \leq \mu([a, b]) \leq b - a + \frac{1}{n}$  for all  $n$ . In particular  $\mu(a) = \mu([a, a]) = a - a = 0$ , it follows that the measure of any finite or countably infinite set is 0.

$\mu((a, b)) = b - a$ , indeed  $[a + \frac{1}{n}, b) \subset (a, b) \subset [a, b]$  for all  $n$ . Hence  $b - a - \frac{1}{n} \leq \mu((a, b)) \leq b - a$  for all  $n$ .

We shall compute the measure of the Cantor set.

The subset we remove from  $[0, 1]$  is the disjoint union  $(\frac{1}{3}, \frac{2}{3}) \cup ((\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})) \cup ((\frac{1}{27}, \frac{2}{27}) \cup (\frac{7}{27}, \frac{8}{27}) \cup (\frac{19}{27}, \frac{20}{27}) \cup (\frac{25}{27}, \frac{26}{27})) \dots$ . It follows that the measure of this set is  $\frac{1}{3} + (\frac{1}{9} + \frac{1}{9}) + (\frac{1}{27} + \frac{1}{27} + \frac{1}{27} + \frac{1}{27}) \dots = \frac{1}{3} + \frac{2}{9} + \frac{4}{27} \dots = \sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n} = 1$ . It follows that the Cantor set has measure  $1 - 1 = 0$ .

### 1.3 Measurable Functions

Let  $f : X \rightarrow Y$  be a function between sets. If  $B \subset Y$  then  $f^{-1}(B) \subset X$  denotes the set of elements  $x \in X$  such that  $f(x) \in B$ . This is called *the inverse image of B under f*

Consider a set  $\Omega$  with a  $\sigma$ -algebra  $\mathcal{F}$  of subsets. A function  $f : \Omega \rightarrow \mathbb{R}$  is said to be measurable (with respect to  $\mathcal{F}$ ) if for any real number  $c$  the set  $\{\omega \in \Omega | f(\omega) < c\} = f^{-1}((-\infty, c))$  belongs to  $\mathcal{F}$ .

Remark that the notion of a function being measurable does not actually involve a measure as defined above but only a  $\sigma$ -algebra. This may seem a bit confusing. The reason is that the  $\sigma$ -algebra is often referred to as the *algebra of measurable sets*. If  $\mathcal{F}$  is the set of all subsets of  $\Omega$  then any function  $f : \Omega \rightarrow \mathbb{R}$  is measurable. If  $\mathcal{F}$  is the  $\sigma$ -algebra consisting of only  $\Omega$  and  $\emptyset$  then the only measurable functions are constant. (*Exercise: Prove this*)

One can show that a function is measurable if and only if  $f^{-1}(B) \in \mathcal{F}$  for any Borel set  $B$ .

If  $f : \Omega \rightarrow \mathbb{R}$  is any function we can consider the  $\sigma$ -algebra generated by all sets of the form  $f^{-1}(B)$  where  $B$  runs through all the Borel sets, we denote this  $\sigma$ -algebra by  $\sigma(f)$ . It is clear that  $f$  is measurable with respect to  $\sigma(f)$  and that it is the smallest  $\sigma$ -algebra with respect to which  $f$  is measurable. Thus  $f$  is measurable if and only if  $\sigma(f) \subset \mathcal{F}$ .

A function  $f : \Omega \rightarrow \mathbb{R}$  is said to be a simple function if  $f$  is measurable and  $f$  only takes countably many values. Let  $\{y_1, y_2, \dots, y_k, \dots\}$  be the values of  $f$  then the sets  $A_k = f^{-1}(\{y_k\})$  are in  $\mathcal{F}$ . If  $f, g$  are simple functions then  $f + g, fg, \lambda f$  where  $\lambda$  is a real number, are also simple functions.

**Theorem 1.3.1** *Let  $f : \Omega \rightarrow \mathbb{R}$  be a function. Then  $f$  is measurable if and only if there exists a sequence of simple functions  $\{f_n\}$  converging uniformly to  $f$*

In layman's terms convergence of a sequence of numbers  $a_n \rightarrow a$  means that  $a_n$  gets closer and closer to  $a$  as  $n$  gets larger and larger. In mathematical terms it means that for all  $\varepsilon > 0$  (think of  $\varepsilon$  as a very small positive number) we can find an  $n_0$  such that when  $m \geq n_0$ ,  $|a_m - a| < \varepsilon$ , in other words all the terms in the sequence from  $n_0$  onwards lie in the open interval  $(a - \varepsilon, a + \varepsilon)$ . For a sequence of functions there are (at least) two kinds of convergence (in fact there are many others). *Uniform convergence* means that for any  $\varepsilon > 0$  we can find  $n_0$  such that for all  $\omega \in \Omega$ ,  $|f_n(\omega) - f(\omega)| < \varepsilon$  i.e. the same  $n_0$  works for all  $\omega$ . By contrast *pointwise convergence* means that  $f_n(\omega) \rightarrow f(\omega)$  for all  $\omega$  but for a given  $\varepsilon$  we might have to choose different  $n_0$ 's for different  $\omega$ 's. Thus uniform convergence implies pointwise convergence but not conversely.

**Example 1.4** *Consider the half-open interval  $(0, 1]$ . Define  $f_n : (0, 1] \rightarrow \mathbb{R}$  by  $f_n(x) = 0$  for  $\frac{1}{n} < x \leq 1$  and  $f_n(x) = 1$  for  $0 < x \leq \frac{1}{n}$ . Then  $\{f_n\}$  converges pointwise in  $(0, 1]$  to the constant function 0 because if  $x \in (0, 1]$  we can choose  $n$  such that  $x > \frac{1}{n}$  then for all  $m \geq n$  we have  $f_m(x) = 0$ . However the convergence is not uniform. Indeed no matter how large an  $n$  we choose, if  $x \in (0, \frac{1}{n}]$ ,  $|f_n(x) - 0| = 1$ .*

Assume then that  $\{f_n\}$  is a sequence of simple functions converging uniformly to  $f$ .

We shall first show that for all real numbers  $c$

$$\{\omega | f(\omega) < c\} = \bigcup_k \bigcup_n \bigcap_{m>n} \{\omega | f_m(\omega) < c - \frac{1}{k}\}$$

This equality will prove the "if" part since the right hand side is obviously in  $\mathcal{F}$ . Assume that  $f(\omega) < c$ . Then we can find  $k$  such that  $f(\omega) < c - \frac{2}{k}$ . Now we can find  $n$  such that when  $m > n$  we have  $|f(\omega) - f_m(\omega)| < \frac{1}{k}$  i.e.  $f(\omega) - \frac{1}{k} < f_m(\omega) < f(\omega) + \frac{1}{k}$ . But then  $f_m(\omega) < f(\omega) + \frac{1}{k} < c - \frac{2}{k} + \frac{1}{k} = c - \frac{1}{k}$  which shows that  $\omega$  is in the set on the right hand side. Next assume  $\omega$  is in the right hand side i.e. there exists  $k, n$  such that for all  $m > n$ ,  $f_m(\omega) < c - \frac{1}{k}$ . Now for all sufficiently large  $m$  we have  $|f(\omega) - f_m(\omega)| < \frac{1}{k}$  but then  $f(\omega) < f_m(\omega) + \frac{1}{k} < c$  i.e.  $\omega$  is in the left hand side. Remark that this part of the proof only uses pointwise convergence and does not use that the  $f_n$ 's

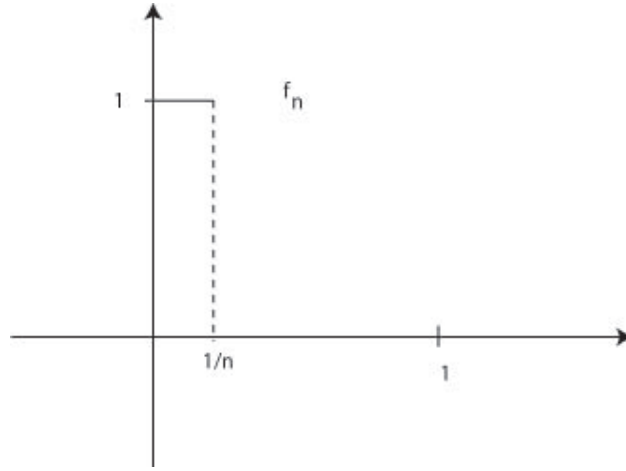
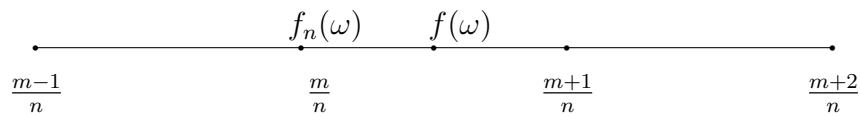


Figure 1: Example 1.4. Pointwise, Non-uniform Convergence

are simple functions. We have in fact shown that if a sequence of measurable functions  $\{f_n\}$  converges pointwise to a function  $f$  then  $f$  is also measurable.

Now assume  $f$  is a measurable function. We have to find a sequence of simple functions  $\{f_n\}$  converging uniformly to  $f$ . This is done as follows: Given an  $n$ , divide the real line into half-open intervals of length  $\frac{1}{n}$  so we mark off the points  $\frac{m}{n}$  where  $m$  runs through the integers. Now for  $\omega \in \Omega$ ,  $f(\omega) \in \mathbb{R}$  and so  $f(\omega)$  lies in precisely one of the half-open intervals  $[\frac{m}{n}, \frac{m+1}{n})$ . We define  $f_n(\omega)$  to be equal to  $\frac{m}{n}$  when  $\frac{m}{n} \leq f(\omega) < \frac{m+1}{n}$



Then we clearly have  $|f_n(\omega) - f(\omega)| < \frac{1}{n}$  for all  $\omega \in \Omega$  and so the sequence of functions  $\{f_n\}$  defined converges uniformly to  $f$  (*Exercise: Verify this using the definition of uniform convergence*). So we just have to verify that the functions  $f_n$  are simple functions. Clearly each  $f_n$  only takes the values  $\{\frac{m}{n}\}$  and so takes only countably many values. It remains to show that they are measurable functions. Let  $c$  be a real number and consider  $\{\omega | f_n(\omega) < c\}$ . Let  $m$  be the unique integer such that  $c \in [\frac{m}{n}, \frac{m+1}{n})$  then  $f_n(\omega) < c$  means

that  $f_n(\omega) < \frac{m+1}{n}$  since  $f_n$  only takes values of the form  $\frac{k}{n}$ ,  $k$  an integer. So  $f_n(\omega)$  is one of the numbers  $\frac{k}{n}$  with  $k < m + 1$ . But  $f_n(\omega) = \frac{k}{n}$  precisely when  $\frac{k}{n} \leq f(\omega) < \frac{k+1}{n}$  so  $f_n(\omega) < \frac{m+1}{n}$  if and only if  $f(\omega) < \frac{m+1}{n}$ . It follows that  $\{\omega | f_n(\omega) < c\} = \{\omega | f_n(\omega) < \frac{m+1}{n}\} = \{\omega | f(\omega) < \frac{m+1}{n}\}$  and the last set is in  $\mathcal{F}$  since  $f$  is measurable.

**Corollary 1.3.1** *If  $f, g : \Omega \rightarrow \mathbb{R}$  are measurable functions then  $f+g, fg, \lambda f$  where  $\lambda$  is a real number are all measurable. If  $f(\omega) \neq 0$  for all  $\omega \in \Omega$  then  $1/f$  is measurable*

Theorem 1.3.1 implies that it is enough to prove this for simple functions. We skip the proof of this.

## 1.4 Integration

Let  $(\Omega, \mathcal{F}, \mu)$  be a measure space i.e.  $\mathcal{F}$  is a  $\sigma$ -algebra and  $\mu$  is a measure on  $\mathcal{F}$ .

Let  $f : \Omega \rightarrow \mathbb{R}$  be a simple function, taking the values  $\{y_k | k = 1, 2, 3, \dots\}$ . Consider the measurable sets (from now on the sets in  $\mathcal{F}$  will be called measurable)  $A_k = \{\omega | f(\omega) = y_k\}$ . We say that  $f$  is integrable over a measurable set  $A$  if the series  $\sum_{k=1}^{\infty} y_k \mu(A_k \cap A)$  is absolutely convergent i.e. if the series of absolute values  $\sum_{k=1}^{\infty} |y_k \mu(A_k \cap A)|$  is convergent.

If  $f$  is a simple function, integrable over  $A$  we define

$$\int_A f d\mu = \sum_{k=1}^{\infty} y_k \mu(A_k \cap A)$$

Remark that absolute convergence of a series implies convergence (Warning: the opposite is false). If  $f, g$  are simple and integrable the same is true for  $f + g, \lambda f$  and  $\int_A (f + g) d\mu = \int_A f d\mu + \int_A g d\mu$  and  $\int_A \lambda f d\mu = \lambda \int_A f d\mu$

Let now  $f$  be an arbitrary measurable function and let  $A$  be a measurable set. We say that  $f$  is integrable over  $A$  if there exists a sequence of simple functions,  $\{f_n\}$ , integrable over  $A$ , converging uniformly to  $f$ . We define

$$\int_A f d\mu = \lim_{n \rightarrow \infty} \int_A f_n d\mu$$

This is called the Lebesgue integral of  $f$  over  $A$  with respect to the measure  $\mu$ .

Implicit in this definition are some statements that need verification: First that the sequence  $\{\int_A f_n d\mu\}$  does in fact have a limit. To prove this we shall show that it is a Cauchy sequence. Intuitively a Cauchy sequence  $\{a_n\}$  is a sequence where the terms come ever closer together, mathematically this is expressed as follows: for all  $\varepsilon > 0$  there is an  $n_0$  such that when  $n, m > n_0$ ,  $|a_n - a_m| < \varepsilon$ . One of the fundamental properties of  $\mathbb{R}$  is that any Cauchy sequence has a limit.

To prove this we need the following result:

**Lemma 1.4.1** *Let  $g$  be a simple function. Assume there exists a constant  $M$  such that  $|g(\omega)| \leq M$  for all  $\omega \in A$ . Then  $g$  is integrable over  $A$  and  $|\int_A g d\mu| \leq M\mu(A)$*

Let  $A_k = \{\omega | g(\omega) = y_k\} = g^{-1}(\{y_k\})$  where  $\{y_k\}_k$  are the countably many values of  $g$ . Then the  $A_k$ 's are measurable and pairwise disjoint and  $A = \bigcup_k A_k \cap A$ . It follows by  $\sigma$ -additivity that  $\mu(A) = \sum_k \mu(A_k \cap A)$ . Hence we have  $\sum_k |y_k \mu(A_k \cap A)| = \sum_k |y_k| \mu(A_k \cap A) \leq M \sum_k \mu(A_k \cap A) = M\mu(A)$ . This shows that  $g$  is integrable. We have

$$|\int_A g d\mu| = |\sum_k y_k \mu(A_k \cap A)| \leq \sum_k |y_k| \mu(A_k \cap A) \leq M\mu(A)$$

which finishes the proof of the lemma.

Now  $f_n \rightarrow f$  uniformly hence for a given  $\varepsilon > 0$  we can find  $n_0$  such that when  $m > n_0$ ,  $|f_m(\omega) - f(\omega)| < \varepsilon/2\mu(A)$  for all  $\omega$ . It follows that for all  $\omega$  and all  $m, n > n_0$  we have

$$\begin{aligned} |f_n(\omega) - f_m(\omega)| &= |f_n(\omega) - f(\omega) + f(\omega) - f_m(\omega)| \\ &\leq |f_n(\omega) - f(\omega)| + |f(\omega) - f_m(\omega)| < \varepsilon/2\mu(A) + \varepsilon/2\mu(A) = \varepsilon/\mu(A) \end{aligned}$$

It then follows from the lemma that for  $n, m > n_0$  we have

$$|\int_A f_n d\mu - \int_A f_m d\mu| = |\int_A (f_n - f_m) d\mu| \leq \frac{\varepsilon}{\mu(A)} \mu(A) = \varepsilon$$

Next we have to verify that  $\int_A f d\mu$  does not depend on the choice of the sequence of simple functions converging to  $f$ . Thus let  $\{g_n\}$  be another sequence of simple functions converging uniformly to  $f$ . We have to show that  $\lim_{n \rightarrow \infty} \int_A g_n d\mu = \lim_{n \rightarrow \infty} \int_A f_n d\mu$ .

Given  $\varepsilon > 0$  we can find  $n_0$  such that for  $n > n_0$ ,  $|f_n(\omega) - f(\omega)| < \varepsilon/2\mu(A)$  and  $|g_n(\omega) - f(\omega)| < \varepsilon/2\mu(A)$  for all  $\omega$ . It follows that  $|g_n(\omega) - f_n(\omega)| \leq |g_n(\omega) - f(\omega)| + |f(\omega) - f_n(\omega)| < \varepsilon/2\mu(A) + \varepsilon/2\mu(A) = \varepsilon/\mu(A)$  for all  $\omega$ . The lemma then gives

$$\left| \int_A g_n d\mu - \int_A f_n d\mu \right| = \left| \int_A (g_n - f_n) d\mu \right| < \frac{\varepsilon}{\mu(A)} \mu(A) = \varepsilon$$

This proves that the terms in the two sequences get closer and closer and so they have the same limit.

The integral  $\int_A f d\mu$  is called the Lebesgue integral of  $f$  over  $A$ . We have the usual statements: if  $f, g$  are integrable over  $A$  then so is  $f + g$  and  $\lambda f$  and  $\int_A (f + g) d\mu = \int_A f d\mu + \int_A g d\mu$ ,  $\int_A \lambda f d\mu = \lambda \int_A f d\mu$ . All of these follow very easily from the corresponding statements for simple functions (*Exercise: Prove these statements*)

## 1.5 Some key facts about integrals

We extend Lemma 1.4.1 to general integrable functions

**Lemma 1.5.1** *Assume  $f$  is measurable and let  $A \subset \Omega$  be a measurable set. Assume  $f$  is bounded on  $A$  i.e. there is a constant  $M$  such that  $|f(\omega)| \leq M$  for all  $\omega \in A$ . Then  $f$  is integrable on  $A$  and  $\int_A f d\mu \leq M\mu(A)$ .*

Let  $\{f_n\}$  be a sequence of simple functions converging uniformly to  $f$ . Let  $\varepsilon > 0$  be given and choose  $n_0$  such that  $|f(\omega) - f_n(\omega)| < \varepsilon/\mu(A)$  for all  $n \geq n_0$  and all  $\omega \in A$ . We have  $|f_n(\omega)| = |f_n(\omega) - f(\omega) + f(\omega)| \leq |f_n(\omega) - f(\omega)| + |f(\omega)| < \varepsilon/\mu(A) + M$ . Hence by Lemma 1.4.1  $f_n$  is integrable on  $A$  and  $\left| \int_A f_n d\mu \right| \leq (M + \varepsilon/\mu(A))\mu(A) = M\mu(A) + \varepsilon$ . Hence  $f$  is also integrable on  $A$  and  $\int_A f d\mu = \lim_{n \rightarrow \infty} \int_A f_n d\mu$ . But then  $\left| \int_A f d\mu \right| = \lim_{n \rightarrow \infty} \left| \int_A f_n d\mu \right| \leq M\mu(A) + \varepsilon$ . Since this holds for all  $\varepsilon > 0$  we get  $\left| \int_A f d\mu \right| \leq M\mu(A)$ .

**Proposition 1.5.1** *Let  $A = \bigcup_i A_i$  be a finite or countable union of pairwise disjoint measurable sets. Assume  $f$  is integrable over  $A$ , then  $f$  is integrable over each  $A_i$  and*

$$\int_A f d\mu = \sum_i \int_{A_i} f d\mu$$



Assume first that  $f$  is a simple function taking the values  $y_1, y_2, y_3, \dots$  and let  $B_k = f^{-1}(\{y_k\}) \cap A$ ,  $B_{ni} = B_k \cap A_i = \{\omega \in A_i | f(\omega) = y_k\}$ .

Then

$$\begin{aligned} \int_A f d\mu &= \sum_k y_k \mu(B_k) = \sum_k y_k \sum_i \mu(B_{ki}) \\ &= \sum_i \sum_k y_k \mu(B_{ki}) = \sum_i \int_{A_i} f d\mu \end{aligned}$$

(here we have used the fact that in an absolutely convergent series, the terms can be rearranged without affecting convergence or the limit)

Next assume  $f$  is integrable over  $A$ . By definition of integrability we can find a sequence of simple functions  $\{f_n\}$  converging uniformly to  $f$ . By the argument above each of the simple functions  $f_n$  are integrable over  $A_i$ . Since  $f_n \rightarrow f$  uniformly on  $A_i$  it follows that  $f$  is integrable on  $A_i$ .

Now let  $\varepsilon > 0$  be given. Choose  $n_0$  such that for  $n > n_0$  and for all  $\omega \in A$ ,  $|f(\omega) - f_n(\omega)| < \varepsilon/2\mu(A)$ . Then

$$\left| \int_{A_i} f d\mu - \int_{A_i} f_n d\mu \right| = \left| \int_{A_i} (f - f_n) d\mu \right| < \frac{\varepsilon}{2\mu(A)} \mu(A_i)$$

Hence we have

$$\sum_i \left| \int_{A_i} f d\mu - \int_{A_i} f_n d\mu \right| < \sum_i \frac{\varepsilon}{2\mu(A)} \mu(A_i) = \frac{\varepsilon}{2\mu(A)} \sum_i \mu(A_i) = \frac{\varepsilon}{2}$$

It follows that

$$\begin{aligned} \left| \int_A f d\mu - \sum_i \int_{A_i} f d\mu \right| &= \left| \int_A f d\mu - \int_A f_n d\mu + \sum_i \int_{A_i} f_n d\mu - \sum_i \int_{A_i} f_n d\mu \right| \\ &\leq \left| \int_A f d\mu - \int_A f_n d\mu \right| + \left| \sum_i \int_{A_i} f_n d\mu - \sum_i \int_{A_i} f d\mu \right| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

where we have used that  $\int_A f_n d\mu = \sum_i \int_{A_i} f_n d\mu$  which we know to hold because  $f_n$  is a simple function. Hence  $\left| \int_A f d\mu - \sum_i \int_{A_i} f d\mu \right| < \varepsilon$  for all positive  $\varepsilon$ . This means that  $\int_A f d\mu = \sum_i \int_{A_i} f d\mu$ .

**Corollary 1.5.1** *If  $f$  is integrable on  $A$  then  $f$  is integrable on every measurable subset  $A' \subset A$*

This follows since  $A = A' \cup (A - A')$ .

**Lemma 1.5.2** *Let  $f$  be a measurable function on  $\Omega$ . If  $A$  is a set of measure 0 then  $f$  is integrable on  $A$  and  $\int_A f d\mu = 0$*

This is very easy to prove with the usual technique. Let  $f_n \rightarrow f$  be a sequence of simple functions converging uniformly to  $f$  on  $A$ . If  $f_n$  takes the values  $y_{n1}, y_{n2}, y_{n3} \dots$  then  $\sum_k y_{nk} \mu(f_n^{-1}(y_{nk}) \cap A)$  is certainly absolutely convergent (all the terms are 0!). Hence  $f_n$  is integrable on  $A$  and  $\int_A f_n d\mu = 0$ . Thus  $f$  is integrable and  $\int_A f d\mu = \lim_{n \rightarrow \infty} \int_A f_n d\mu = 0$ .

**Lemma 1.5.3** *Let  $f$  be non-negative except for a subset of measure 0 on  $A$  (if a property holds outside a set of measure 0 we shall say it holds almost everywhere, abbreviated a.e.) i.e.  $f(\omega) \geq 0$  a.e on  $A$ . Then  $\int_A f d\mu \geq 0$*

Since the integral doesn't care about a subset of measure 0 we might as well assume that  $f$  is non-negative on all of  $A$ .

Let  $A'$  be the set  $\{\omega \in A | f(\omega) > 0\}$  so  $f$  is 0 on  $A - A'$ . Then  $\int_A f d\mu = \int_{A'} f d\mu + \int_{A-A'} f d\mu = \int_{A'} f d\mu$ .

Consider the sets  $A_n = \{\omega \in A' | f(\omega) > 1/n\}$ , then we have  $\bigcup A_n = A'$  and  $\dots A_{n-1} \subset A_n \subset A_{n+1} \subset \dots \subset A'$ . We have  $\int_{A'} f d\mu = \int_{A_n} f d\mu + \int_{A'-A_n} f d\mu$  and  $A' - A_n = \{\omega \in A' | f(\omega) \leq 1/n\}$ . It follows that  $|\int_{A'-A_n} f d\mu| \leq \frac{1}{n} \mu(A' - A_n) \leq \frac{1}{n} \mu(A') \rightarrow 0$  for  $n \rightarrow \infty$  and hence  $\int_{A_n} f d\mu \rightarrow \int_{A'} f d\mu$ .

Let  $\{g_k\}$  be a sequence of simple functions converging uniformly to  $f$ . Choose  $k_0$  so that for  $k > k_0$   $|f(\omega) - g_k(\omega)| < 1/n$  for all  $\omega$  in  $A$ .

Since  $f(\omega) > 1/n$  for  $\omega \in A_n$  and  $g_k(\omega) \in (f(\omega) - \frac{1}{n}, f(\omega) + \frac{1}{n})$  we must have  $g_k(\omega) > 0$  on  $A_n$ . Let  $g_k$  take the values  $x_1, x_2, x_3, \dots$  on  $A_n$ , then  $x_i > 0$  for all  $i$ .

It follows that  $\int_{A_n} g_k d\mu = \sum_i x_i \mu(\{\omega \in A_n | g_k(\omega) = x_i\}) > 0$  but then  $\int_{A_n} f d\mu = \lim_{k \rightarrow \infty} \int_{A_n} g_k d\mu \geq 0$ . Now  $\int_A f d\mu = \int_{A'} f d\mu = \lim_{n \rightarrow \infty} \int_{A_n} f d\mu \geq 0$ .

**Corollary 1.5.2** *Let  $f$  and  $g$  be integrable on  $A$  and assume  $f \geq g$  a.e. then  $\int_A f d\mu \geq \int_A g d\mu$*

This follows by taking  $f - g$  which is non-negative a.e. and applying Lemma 1.5.2

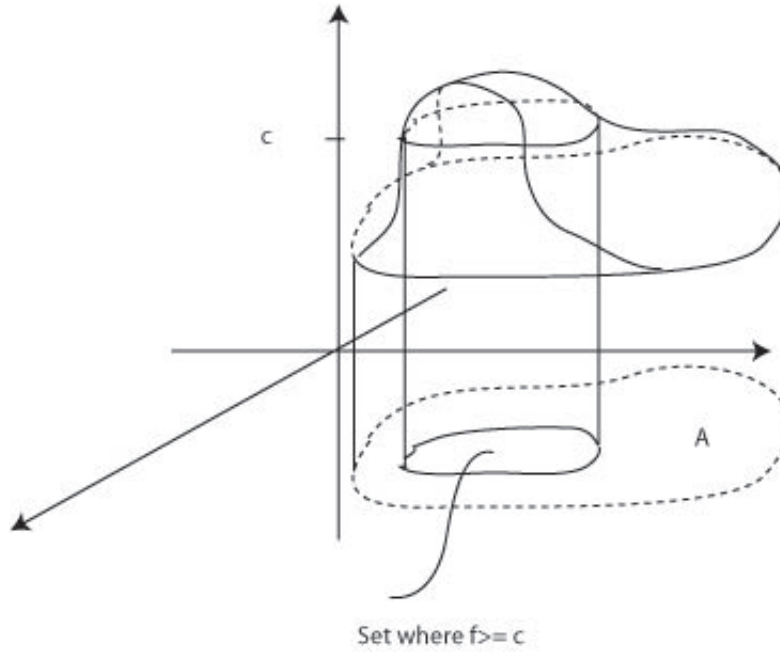


Figure 2: Chebyshev's Inequality

**Theorem 1.5.1** (*Chebyshev's inequality*) Let  $f$  be non-negative and integrable on  $A$  and let  $c > 0$ . Then  $\mu(\{\omega \in A \mid f(\omega) \geq c\}) \leq \frac{1}{c} \int_A f d\mu$

Let  $A' = \{\omega \in A \mid f(\omega) \geq c\}$  then  $\int_A f d\mu = \int_{A'} f d\mu + \int_{A-A'} f d\mu \geq \int_{A'} f d\mu$  because  $\int_{A-A'} f d\mu \geq 0$  as  $f$  is a non-negative function ( Lemma 1.5.2). On  $A'$   $f - c$  is a non-negative function so we have

$$0 \leq \int_{A'} (f - c) d\mu = \int_{A'} f d\mu - \int_{A'} c d\mu = \int_{A'} f d\mu - c\mu(A')$$

Thus  $\int_A f d\mu \geq \int_{A'} f d\mu \geq c\mu(A')$ .

**Corollary 1.5.3** If  $\int_A |f| d\mu = 0$ , then  $f$  is identically 0 a.e.

By Chebyshev's inequality we have  $\mu(\{\omega \in A \mid |f(\omega)| \geq 1/n\}) \leq n \int_A |f| d\mu = 0$ . Since  $\{\omega \in A \mid |f(\omega)| > 0\} = \bigcup \{\omega \in A \mid |f(\omega)| \geq 1/n\}$  we have  $\mu(\{\omega \in A \mid |f(\omega)| > 0\}) \leq \sum^n \mu(\{\omega \in A \mid |f(\omega)| \geq 1/n\}) = 0$  (*Exercise: Show that a countable union of sets of measure 0 is a set of measure 0*).

Assume now that  $f : \Omega \rightarrow \mathbb{R}$  is a non-negative integrable function. It follows from the previous results that the function on the  $\sigma$ -algebra  $\mathcal{F}$  given by  $A \mapsto \int_A f d\mu$  defines a measure,  $\mu_f$ , on  $\mathcal{F}$ . Indeed Lemma 1.5.1 shows  $\mu_f(A) \geq 0$  and Proposition 1.5.1 proves the  $\sigma$ -additivity.

**Definition 1.5.1** *Let  $\mu$  and  $\nu$  be measures on  $\mathcal{F}$ . Then  $\nu$  is said to be absolutely continuous with respect to  $\mu$  if  $\mu(A) = 0$  implies  $\nu(A) = 0$ .*

As a consequence of Lemma 1.5.2 we see that if  $f$  is a non-negative integrable function on  $\Omega$  then the measure  $\mu_f$  is absolutely continuous with respect to  $\mu$ .

There is a very important theorem which states that every measure which is absolutely continuous with respect to  $\mu$  is of the form  $\mu_f$  for some non-negative integrable function,

**Theorem 1.5.2** (*Radon-Nikodym*) *Let  $\mu$  and  $\nu$  be measures on  $\mathcal{F}$  and assume that  $\nu$  is absolutely continuous with respect to  $\mu$ . Then there exists a non-negative, integrable function  $f$  such that  $\nu = \mu_f$  i.e. for all  $A \in \mathcal{F}$  we have  $\nu(A) = \int_A f d\mu$ . The function  $f$  is uniquely determined a.e. (i.e. if  $\mu_g = \nu = \mu_f$  then  $f = g$  a.e.)*

We shall only sketch the proof but even before we can do that we need another important result

**Theorem 1.5.3** (*Levi*) *Let  $\{f_k\}$  be a sequence of functions, integrable on a measurable set  $A$ . Assume that  $f_1(\omega) \leq f_2(\omega) \leq f_3(\omega) \leq \dots \leq f_k(\omega) \leq \dots$  for all  $\omega \in A$ . Assume further that there is a constant  $M$  such that  $\int_A f_k d\mu \leq M$  for all  $k$ . Then the sequence  $\{f_k(\omega)\}$  is convergent a.e. on  $A$  and if we define  $f(\omega) = \lim_{k \rightarrow \infty} f_k(\omega)$  if the limit exists and  $f(\omega) = 0$  if the limit does not exist (this can happen only on a set of measure 0), then  $f$  is integrable on  $A$  and  $\int_A f d\mu = \lim_{k \rightarrow \infty} \int_A f_k d\mu$*

The intuition behind Levi's theorem is that if there is a set  $B \subset A$  of positive measure such that  $f_k(\omega) \rightarrow \infty$  on  $B$  then the sequence of integrals  $\int_B f_k d\mu$  will also tend to  $\infty$  and hence  $\int_A f_k d\mu = \int_B f_k d\mu + \int_{A-B} f_k d\mu$  can't be bounded.

Let  $g_k = f_k - f_1$  then the sequence  $\{g_k(\omega)\}$  is also increasing and the  $g_k$ 's are non-negative functions. Clearly they are also integrable on  $A$  and

$\int_A g_k d\mu = \int_A f_k d\mu - \int_A f_k d\mu \leq M - \int_A f_1 d\mu$ . If we can prove the theorem for the sequence  $\{g_k\}$  then it will follow for the  $f_k$ 's (*Exercise: Show this*) so we may as well assume to begin with that the  $f_k$ 's are non-negative functions on  $A$ . Let  $B$  be the set  $\{\omega \in A | f_k(\omega) \rightarrow \infty\}$  and  $B_{r,k} = \{\omega \in A | f_k(\omega) \geq r\}$ . Then we have  $B = \bigcap_r \bigcup_k B_{r,k}$  (*Exercise: Show this*). By Chebyshev's inequality (which we can apply since the  $f_k$ 's are now assumed to be non-negative) we have  $\mu(B_{r,k}) \leq 1/r \int_A f_k d\mu \leq M/r$  for all  $r$  and  $k$ . Since the  $f_k$ 's form an increasing sequence we have  $\dots B_{r,k-1} \subset B_{r,k} \subset B_{r,k+1} \subset \dots$ . Let  $B_r = \bigcup_k B_{r,k}$  and put  $C_k = B_r - B_{r,k}$  then the  $C_k = (C_k - C_{k+1}) \cup (C_{k+1} - C_{k+2}) \cup \dots$ . This is a disjoint union hence by  $\sigma$ -additivity we have  $\mu(C_k) = \sum_{i \geq k} \mu(C_i - C_{i+1})$ . In particular the series  $\sum_{i=1} \mu(C_i - C_{i+1}) = \mu(C_1)$  is convergent and so the sequence of tails  $\{\sum_{i \geq k} \mu(C_i - C_{i+1})\}$  tend to 0. It follows that  $\mu(C_k) \rightarrow 0$  or equivalently  $\mu(B_r) - \mu(B_{r,k}) \rightarrow 0$ . Since  $\mu(B_{r,k}) \leq M/r$  it follows that  $\mu(B_r) \leq M/r$ . But  $B \subset B_r$  for all  $r$  and so  $\mu(B) \leq M/r$  for all  $r$  which means that  $\mu(B) = 0$ .

Since  $\{f_k(\omega)\}$  is an increasing sequence, it is either bounded or it goes to  $\infty$ . Thus outside the measure 0 set  $B$  the sequence is bounded. A fundamental property of the real numbers is that any non-empty subset  $S \subset \mathbb{R}$  which is bounded above has a smallest upper bound (it is denoted  $\sup S$ ). For  $\omega \in A - B$  we put  $f(\omega) = \sup \{f_k(\omega)\}$ . We claim that  $f_k(\omega) \rightarrow f(\omega)$ . Let  $\varepsilon > 0$  be given then since  $f(\omega)$  is the least upper bound  $f(\omega) - \varepsilon$  is not an upper bound. Hence there is an  $n_0$  such that  $f(\omega) - \varepsilon < f_{n_0}(\omega) \leq f(\omega)$ . Since the sequence of  $f_k(\omega)$ 's is increasing, we have  $f(\omega) - \varepsilon < f_{n_0}(\omega) \leq f_k(\omega) \leq f(\omega)$  for all  $k \geq n_0$  hence  $|f(\omega) - f_k(\omega)| < \varepsilon$  for  $k \geq n_0$ . We skip the rest of the proof which relies on a theorem known as Lebesgue's bounded convergence theorem. Levi's theorem is also known as the monotone convergence theorem.

We now return to the proof of the Radon-Nikodym theorem.

Let  $K$  be the set of all  $\mu$ -integrable functions on  $\Omega$  such that  $\int_A \phi d\mu \leq \nu(A)$  for all  $A \in \mathcal{F}$ ,  $K$  is not empty since the constant function 0 is in  $K$ . Let  $M = \sup_{\phi \in K} \{\int_{\Omega} \phi d\mu\}$ . Remark that the  $\sup$  exists because  $\int_{\Omega} \phi d\mu$  is bounded by  $\nu(\Omega)$  (here it is important that  $\nu$  takes values in  $\mathbb{R}$  i.e. we do not allow  $\nu$  to take the value  $\infty$ ).

For each  $n$  consider  $M - 1/n$ , since  $M - 1/n < M$  this is not an upper bound and so we can find  $f_n \in K$  such that  $M - 1/n < \int_{\Omega} f_n d\mu \leq M$ . It

follows that the sequence  $\{\int_{\Omega} f_n d\mu\}$  converges to  $M$ .

Next define  $g_n(\omega) = \max\{f_1(\omega), f_2(\omega), \dots, f_n(\omega)\}$ , for all  $\omega \in \Omega$ . It is clear that  $g_1(\omega) \leq g_2(\omega) \leq \dots \leq g_k(\omega) \leq \dots$ .

For  $A \in \mathcal{F}$  consider  $A_k = \{\omega \in A | g_n(\omega) = f_k(\omega)\}$  i.e.  $A_k$  is the set of elements  $\omega \in A$  where  $f_k(\omega)$  is the largest of the numbers  $\{f_1(\omega), \dots, f_n(\omega)\}$ . Let  $B_1 = A_1, B_2 = A_2 - A_1, A_3 = (A_1 \cup A_2), \dots, B_n = A_n - (A_1 \cup A_2 \cup \dots \cup A_{n-1}), \dots$ . Then the  $B_i$ 's are pairwise disjoint and  $A = \bigcup_i B_i$  (*Exercise: Prove this*).

It follows that  $\int_A g_n d\mu = \sum_k \int_{B_k} g_n d\mu = \sum_k \int_{B_k} f_k d\mu \leq \sum_k \nu(B_k) = \nu(A)$ . This shows that  $g_n \in K$ . We have  $\int_{\Omega} f_n d\mu \leq \int_{\Omega} g_n d\mu \leq M$  where the first inequality follows since  $f_n \leq g_n$  (and using Corollary 1.5.2) and the second because we have shown  $g_n \in K$ . Since  $\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu \rightarrow M$  it follows that  $\int_{\Omega} g_n d\mu \rightarrow M$ .

It now follows from Levi's theorem that  $\{g_n(\omega)\}$  is convergent a.e. and we put  $f(\omega) = \lim g_n(\omega)$  if  $\{g_n(\omega)\}$  is convergent and  $f(\omega) = 0$  if not. Levi's theorem also shows that  $f$  is integrable and  $\int_A f d\mu = \lim \int_A g_n d\mu \leq \nu(A)$  hence  $f \in K$ . To show that  $f$  is the function we are looking for, one looks at  $\nu(A) - \int_A f d\mu$ . Since this is non-negative, it is in fact a measure on  $\mathcal{F}$  and the final step (which we shall skip) is to show that this measure is identically 0.

The function  $f$  whose existence is asserted by the theorem is called the Radon-Nikodym derivative of  $\nu$  with respect to  $\mu$  and is denoted

$$f = \frac{d\nu}{d\mu}$$

Example: Let  $\Omega$  be a finite set and assume  $\mu(\omega)$  and  $\nu(\omega)$  are positive for all  $\omega \in \Omega$ . Then  $\frac{d\nu}{d\mu}(\omega) = \frac{\nu(\omega)}{\mu(\omega)}$  for all  $\omega \in \Omega$ . To see this, let  $A \subset \Omega$  then

$$\int_A \frac{d\nu}{d\mu} d\nu = \sum_{\omega \in A} \frac{d\nu}{d\mu}(\omega) \nu(\omega) = \sum_{\omega \in A} \frac{\mu(\omega)}{\nu(\omega)} \nu(\omega) = \sum_{\omega \in A} \mu(\omega) = \mu(A)$$

The importance of the Radon-Nikodym theorem is that it allows us to easily change measures in the integral. This is analogous to the usual formulas for changing variables in the Riemann integral.

**Theorem 1.5.4** *Let  $g : \Omega \rightarrow \mathbb{R}$  be a function and let  $\nu$  be a measure, absolutely continuous with respect to  $\mu$ . Then  $g$  is integrable with respect to  $\nu$  if and only if  $g \frac{d\nu}{d\mu}$  is integrable with respect to  $\mu$  and in this case we have*

for all  $A \in \mathcal{F}$

$$\int_A g d\nu = \int_A g \frac{d\nu}{d\mu}$$

First we prove it for a simple non-negative function. Thus let  $g$  be a simple function taking the values  $\{c_i \geq 0\}_{i=1,2,\dots}$  and let  $A_i = g^{-1}(c_i)$ . Then  $g$  is integrable over  $A$ ,  $A \in \mathcal{F}$ , with respect to  $\nu$  if and only if

$$\sum c_i \nu(A_i \cap A) < \infty$$

We have  $\int_A g d\nu = \sum c_i \nu(A \cap A_i) = \sum c_i \int_{A_i \cap A} \frac{d\nu}{d\mu} d\mu = \int_A g \frac{d\nu}{d\mu} d\mu$ , which proves the change of measure formula. For a general non-negative function  $g$  we write it as a uniform limit of integrable, non-negative, simple functions and pass to the limit in the integrals. Finally for general functions we apply the following often used trick: consider  $g^+ = \max(g, 0)$  and  $g^- = \max(-g, 0)$ . Then  $g^+$  and  $g^-$  are non-negative functions and  $g = g^+ - g^-$  and we can apply the argument above to  $g^+$  and  $g^-$  separately.

## 2 Probability Theory

### 2.1 Probability Spaces and Random Variables

**Definition 2.1.1** A measure space  $(\Omega, \mathcal{F}, \mu)$  is called a probability space if  $\mu(\Omega) = 1$ .

**Definition 2.1.2** Let  $(\Omega, \mathcal{F}, \mu)$  be a probability space. A random variable  $X$  is a measurable function  $X : \Omega \rightarrow \mathbb{R}$ .

The sets in  $\mathcal{F}$  are known as events. If  $A$  is an event then  $\mu(A)$  is called the probability of the event  $A$ . Remark that for all events  $0 \leq \mu(A) \leq 1$ .

**Definition 2.1.3** Let  $X$  be a random variable and assume that  $X$  is integrable on  $\Omega$  then the expectation of  $X$  is defined by  $\mathbb{E}(X) = \int_{\Omega} X d\mu$ .

Example: Consider the set,  $\Omega_n$  of outcomes of tossing a coin  $n$  times. We shall denote heads by 1 and tails by 0. Thus  $\Omega_n$  is the set of vectors of length  $n$  with only 0's and 1's and  $\Omega_n$  is a finite set with  $2^n$  elements. For instance if  $n = 3$ ,  $\Omega_3$  consists of the 8 elements:

$$\Omega_3 = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$$

Let  $\eta \in \Omega_i$  and let

$$U_\eta = \{\omega = (a_1, a_2, \dots, a_n) \in \Omega_n \mid (a_1, a_2, \dots, a_i) = \eta\}$$

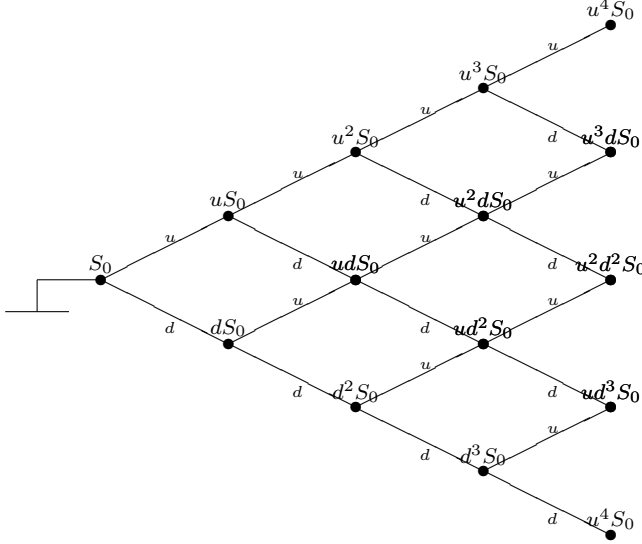
For each  $i \leq n$  let  $\mathcal{S}_i$  be the set of all subsets of the form  $U_\eta$  where  $\eta$  runs through  $\Omega_i$ . Let  $\mathcal{F}_i$  be the  $\sigma$ -algebra generated by  $\mathcal{S}_i$ . Clearly  $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n$ . Indeed if  $\eta \in \Omega_i$  say  $\eta = (a_1, a_2, \dots, a_i)$  we put  $\eta_0 = (a_1, a_2, \dots, a_i, 0)$  and  $\eta_1 = (a_1, a_2, \dots, a_i, 1)$ . Then  $U_\eta = U_{\eta_0} \cup U_{\eta_1} \in \mathcal{F}_{i+1}$ . We have  $\mathcal{F}_0 = \{\Omega_n, \emptyset\}$  and  $\mathcal{F}_n = \mathcal{P}(\Omega_n)$ . We define a probability measure on  $\mathcal{F}_n$  by setting  $\mu(\{a\}) = 1/2^n$  for all elements  $a \in \Omega_n$  and extending it by additivity to all of  $\mathcal{P}(\Omega)$ . We have  $\mu(U_\eta) = 2^{-i}$

We shall use this very simple example to study a problem in finance.

Consider the evolution of the price of a stock over  $n$  time periods. If the price of the stock at time  $t = i$  is  $S_i$  then the price at time  $t = i + 1$  is either  $S_{i+1} = uS_i$  or  $S_{i+1} = dS_i$  where  $0 < d < 1 < u$ . If the price of the stock at time  $t = 0$  is  $S_0$  then the possible stock prices at time  $t = i$  are  $u^i S_0, u^{i-1} d S_0, \dots, u^{i-k} d^k S_0, \dots, d^i S_0$ . We can visualize this using a binomial tree



## Binomial tree of stock prices for $n = 4$ periods



A path through the tree can be described by an element of  $\Omega$  by letting 1 correspond to an up move and 0 to a down move, for example in the tree above the element  $(0, 1, 1, 0)$  would correspond to the path

$$(S_0, dS_0, udS_0, u^2 dS_0, u^2 d^2 S_0)$$

We let  $S_i$  denote the function whose value on an element  $\omega \in \Omega_n$  is the stock price at time  $t = i$  on the path described by  $\omega$ . We notice that for each  $i \leq n$ ,  $S_i$  is measurable with respect to the  $\sigma$ -algebra  $\mathcal{F}_i$  so  $S_i$  is a random variable with respect to  $\mathcal{F}_i$ . Indeed  $\{\omega | S_i(\omega) = u^{i-k} d^k S_0\} = \bigcup U_\eta$  where  $\eta$  runs over the  $\binom{i}{k}$  elements in  $\Omega_i$  with precisely  $k$  of its coordinates = 0. Hence  $S_i^{-1}(\{u^{i-k} d^k S_0\}) \in \mathcal{F}_i$ . Notice also that  $S_i$  is not a random variable with respect to  $\mathcal{F}_{i-1}$ , for instance  $S_1$  is not constant hence is not measurable with respect to  $\mathcal{F}_0$ . The  $\sigma$ -algebra  $\mathcal{F}_i$  can be interpreted as the information that is revealed to us at time  $t = i$ . Indeed if the stock price follows the path given by an unknown element  $\omega \in \Omega_n$ , then knowing the stock prices  $(S_0, S_1, \dots, S_i)$  reveals the first  $i$  coordinates of  $\omega$ . In other words we know that  $\omega \in U_\eta$  where  $\eta$  is determined by the stock prices at times  $t = 1, \dots, i$ . In the next time step the stock price  $S_{i+1}$  reveals the next coordinate and hence tells us which set in  $\mathcal{F}_{i+1}$ ,  $\omega$  lies in.

The expectation  $\mathbb{E}(S_i)$  is computed as follows: if  $S_i(\omega) = u^{i-k} d^k S_0$ , then there are precisely  $k$  0's in the first  $i$  coordinates of  $\omega$ . Now the number of

elements in  $\Omega_n$  with  $k$  0's among the first  $i$  coordinates is  $\binom{i}{k}2^{n-i}$  and each have probability  $1/2^n$ . It follows that the event that  $\omega$  has  $k$  0's among its first  $i$  coordinates has probability  $\binom{i}{k}2^{-i}$ . Then  $\mathbb{E}(S_i) = \sum_{k=0}^i u^{i-k}d^k S_0 \binom{i}{k}2^{-i} = 2^{-i}S_0 \sum_{k=0}^i \binom{i}{k}u^{i-k}d^k = (\frac{u+d}{2})^i S_0$ .

Let  $\eta \in \Omega_i$  and consider the set  $U_\eta \in \mathcal{F}_i$ . Remark that  $S_i$  is constant on  $U_\eta$ , its value is  $u^{i-j}d^j S_0$ , where  $j \leq i$  is the number of 0's in  $\eta$ . Consider  $\int_{U_\eta} S_{i+1} d\mu$ .  $S_{i+1}$  takes the two values  $u(u^{i-j}d^j S_0) = uS_i$  and  $d(u^{i-j}d^j S_0) = dS_i$  on  $U_\eta$ . The probability of  $U_\eta$  is  $2^{-i}$  and hence the probability of  $\{S_{i+1} = uS_i\}$  is  $2^{-i-1}$ . Thus we get  $\int_{U_\eta} S_{i+1} d\mu = u2^{-i-1}S_i + d2^{-i-1}S_i = \frac{(u+d)}{2}2^{-i}S_i = \int_{U_\eta} \frac{(u+d)}{2}S_i d\mu$ . It follows that for all  $A \in \mathcal{F}_i$  we have  $\int_A S_{i+1} d\mu = \int_A \frac{u+d}{2}S_i d\mu$ .

Consider now a european call option on the stock with expiration at time  $t = n$  and strike price  $K$ . What is the fair price of this option at time  $t = i$ ,  $i = 0, 1, 2, \dots, n$ ? Let  $C_i$  denote the function on  $\Omega_n$  whose value on an element  $\omega$  is the price of the option at time  $t = i$  on the path described by  $\omega$ . Assume that  $u^{n-m}d^m S_0 < K \leq u^{n-m+1}d^{m-1}S_0$ , then the price of the option at expiration  $t = n$  is  $u^{n-j}d^j S_0 - K$  if  $j \leq m - 1$  and 0 otherwise. Thus  $C_n = \max(S_n - K, 0)$ .

Let  $r$  be the risk free interest rate at which we can invest and borrow i.e. \$1 invested at time  $t = j$  is guaranteed to pay off  $\$(1+r)^{i-j}$  at time  $t = i$ . To price the option we shall use the no arbitrage principle which states that the price of an investment which with certainty pays \$ $a$  after  $i$  time periods is  $\$(1+r)^{-i}a$ . In other words a risk free investment yields the risk free return.

Assume we know the values of  $C_{i+1}$  we shall first show that we can then compute the values of  $C_i$ . Consider a portfolio consisting of  $\Delta$  shares of the stock, a risk free investment of \$ $B$  (where  $B$  can be positive or negative, a negative  $B$  means we are borrowing) and short one option. If we are at time  $t = i$  then at the node in the stock price tree determined by  $i$  and an element  $\omega \in \Omega_n$  the value of this portfolio is  $\Delta S_i(\omega) + B - C_i(\omega)$ . As we move one step forward in time the stock price can move either up or down, if we let  $C_u$  (resp.  $C_d$ ) denote the value of the option at time  $t = i + 1$  after an upward move (resp. after a downward move) the value of the portfolio is either  $\Delta uS_i + (1+r)B - C_u$  (if we move up) or  $\Delta dS_i + (1+r)B - C_d$  (if we move down). We shall find  $\Delta$  and  $B$  so that the value of the portfolio is 0 in both cases. This is simply solving two linear equations with two unknowns

and we find

$$\Delta = \frac{C_u - C_d}{(u - d)S_i}, \quad B = \frac{1}{1 + r} \left( \frac{uC_d - dC_u}{u - d} \right)$$

Hence we get

$$C_i(\omega) = \frac{C_u - C_d}{u - d} + \frac{1}{r + 1} \frac{uC_d - dC_u}{u - d} = \frac{1}{1 + r} \left( \frac{(1 + r) - d}{u - d} C_u + \frac{u - (1 + r)}{u - d} C_d \right)$$

Let  $p = \frac{(1+r)-d}{u-d}$  then we have  $1 - p = \frac{u-(1+r)}{u-d}$ . It is reasonable to assume that  $d < 1 + r < u$ , indeed if  $u \leq 1 + r$  then a risk free return would yield at least as much as a risky investment and so nobody would invest in the stock. Similarly if  $1 + r \leq d$  the risky return would always yield as much as the riskfree investment and with a chance for a strictly bigger return, thus nobody would invest in the riskfree asset. Under this assumption we have  $0 < p < 1$  and so we can view  $p$  as a probability. It follows that the value of the option  $C_i(\omega)$  is the discounted expectation with respect to  $p$  of the value of the option at time  $t = i + 1$ .

Now we can continue this way  $C_u = \frac{1}{1+r}(pC_{uu} + (1-p)C_{ud})$  and  $C_d = \frac{1}{1+r}(pC_{du} + (1-p)C_{dd})$  so  $C_i(\omega) = \frac{1}{(1+r)^2}(p^2C_{uu} + 2p(1-p)C_{ud} + (1-p)^2C_{dd})$ , the  $C_{uu}$  etc. are option values at time  $t = i + 2$ , remark that here we have used that  $C_{ud} = C_{du}$  i.e. that the value of the option only depends on the node in the tree, but this is true at expiration  $t = n$  and so again using the formula it holds at  $t = n - 1$  etc. Continuing this way we get that  $C_i(\omega) = \frac{1}{(1+r)^{n-i}} \sum_{j=0}^{n-i} \binom{n-i}{j} p^{n-i-j} (1-p)^j C_{uu\dots dd}$  where  $uu\dots dd$  is a sequence with  $n - i - j$   $u$ 's and  $j$   $d$ 's. Now the  $C_{uu\dots dd}$  are the option values at expiration and those we know namely  $C_{uu\dots dd} = \max(u^{n-i-j} d^j S_i(\omega) - K, 0)$  and so we get the formula

$$C_i(\omega) = \frac{1}{(1+r)^{n-i}} \sum_{j=0}^{n-i} \binom{n-i}{j} p^{n-i-j} (1-p)^j \max(u^{n-i-j} d^j S_i(\omega) - K, 0)$$

It follows from this formula that  $C_i$  is a random variable with respect to the  $\sigma$ -algebra  $\mathcal{F}_i$ .

We have  $\max(u^{n-j}d^j S_0 - K, 0) = u^{n-j}d^j S_0 - K$  for  $j \leq m - 1$  and 0 otherwise so we get the following formula for the option price at  $t = 0$ :

$$C_0 = \left( \frac{1}{(1+r)^n} \sum_{j=0}^{j=m-1} \binom{n}{j} p^{n-j}(1-p)^j u^{n-j} d^j \right) S_0 - \left( \frac{1}{(1+r)^n} \sum_{j=0}^{j=m-1} \binom{n}{j} p^{n-j}(1-p)^j \right) K$$

As we remarked above we have  $0 < p < 1$  and so we can define a probability measure on  $\Omega_n$  by  $p(\omega) = p^j(1-p)^{n-j}$  where  $j$  is the number of 1's in  $\omega$ . Let  $\eta \in \Omega_i$  and consider the set  $U_\eta$ . Then  $C_i$  is constant on this set (since  $S_i$  is constant on  $U_\eta$ ).  $C_{i+1}$  takes the values  $C_u$  and  $C_d$  (using the previous notation) and so we get  $\int_{U_\eta} C_{i+1} dp = p(U_\eta)(pC_u + (1-p)C_d)$  and hence  $\int_{U_\eta} C_i dp = \frac{1}{1+r} \int_{U_\eta} C_{i+1} dp$ . Let  $M_i$  denote the random variable  $\frac{1}{(1+r)^i} C_i$  then  $M_i$  is measurable with respect to  $\mathcal{F}_i$  and we have  $\int_A M_i dp = \int_A M_{i+1} dp$  for all  $A \in \mathcal{F}_i$ . This relation says that  $\{M_0, M_1, \dots, M_n\}$  is a *martingale* with respect to the sequence of  $\sigma$ -algebras  $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n$ . We shall return to the notion of martingale later on.

## 2.2 Conditional Probability and Independence

Let  $(\Omega, \mathcal{F}, \mu)$  be a probability space.

**Definition 2.2.1** Let  $A, B \in \mathcal{F}$  be two events. We define the probability of  $A$  conditional on  $B$  by  $\mu(A|B) = \mu(A \cap B)/\mu(B)$

**Proposition 2.2.1** Let  $\{B_i\}$  be a countable set of pairwise disjoint events such that  $\Omega = \bigcup_i B_i$ . If  $A$  is any event then  $\mu(A) = \sum_i \mu(A|B_i)\mu(B_i)$

We have  $A = \bigcup_i A \cap B_i$  and hence  $\mu(A) = \sum_i \mu(A \cap B_i) = \sum_i \mu(A|B_i)\mu(B_i)$

**Definition 2.2.2** Two events  $A, B \in \mathcal{F}$  are said to be independent if  $\mu(A|B) = \mu(A)$  or equivalently if  $\mu(A \cap B) = \mu(A)\mu(B)$

**Example 2.1** Consider the probability space  $\Omega_2$ . Let  $A = \{(1, 1), (1, 0)\}$  and  $B = \{(1, 0), (0, 1)\}$ . In words  $A$  is the event "heads" on the first toss and  $B$  is the event that the two tosses are different. If we know that the first toss is heads then the outcome is in  $B$  if and only if the second toss

is tails, which has probability  $1/2 = \mu(B)$ , so intuitively  $\mu(B|A) = \mu(B)$ . We have  $A \cap B = \{(1, 0)\}$  and  $\mu(A) = \mu(B) = 1/2$  and  $\mu(A \cap B) = 1/4$ . Thus  $A$  and  $B$  are independent in the formal sense as well as intuitively. If we change the probability measure this is no longer true, let us for instance assume that the probability of 1 is  $p = .001$  so the probability of 0 is .999 then  $(1, 1)$  is very unlikely, it has probability  $p^2 = .000001$  thus knowing that the first toss is a 1 makes it very unlikely that the second toss is also a 1 and so intuitively under this probability measure the two events would not be independent. In fact  $p(A \cap B) = .000999$  and  $p(A) = .001$ , so  $p(B|A) = .999$ , but  $p(B) = .001998$

**Definition 2.2.3** Let  $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}$  be sub- $\sigma$ -algebras. Then  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are said to be independent if for every pair of events  $A \in \mathcal{F}_1$ ,  $B \in \mathcal{F}_2$ ,  $A$  and  $B$  are independent

**Example 2.2** Consider again  $\Omega_2$  and the  $\sigma$ -algebra  $\mathcal{F}_1$  i.e. the sigma-algebra generated by the sets  $U_\eta$  where  $\eta$  runs through  $\Omega_1 = \{0, 1\}$ . Then  $\mathcal{F}_1$  consists of the sets  $\emptyset, \Omega_2, \{(0, 0), (0, 1)\}, \{(1, 0), 1, 1\}$ . We can also consider sets of the form  $U^\eta = \{\omega | (\omega_{n-i}, \dots, \omega_n) = \eta\}$  where  $\eta$  runs through  $\Omega_i$ . Let  $\mathcal{F}^1$  be the sigma-algebra generated by the  $U^\eta$ 's where  $\eta$  runs through  $\Omega_1$  then  $\mathcal{F}^1$  consists of the sets  $\emptyset, \Omega_2, \{(0, 0), (1, 0)\}, \{(0, 1), (1, 1)\}$ . These two  $\sigma$ -algebras are independent

**Definition 2.2.4** Two random variables  $X, Y : \Omega \rightarrow \mathbb{R}$  are said to be independent if the  $\sigma$ -algebras they generate,  $\sigma(X)$  and  $\sigma(Y)$  are independent

**Definition 2.2.5** The variance of a random variable  $X : \Omega \rightarrow \mathbb{R}$  is

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2 + \mathbb{E}(X)^2 - 2X\mathbb{E}(X)) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

If  $X, Y : \Omega \rightarrow \mathbb{R}$  are random variables the covariance is defined by

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

The correlation coefficient is defined by

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

**Proposition 2.2.2** *Let  $X_1, X_2, \dots, X_n$  be random variables. Then*

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n) + \sum_{i \neq j} \text{Cov}(X_i, X_j)$$

We have

$$\mathbb{E}((X_1 + X_2 + \dots + X_n)^2) = \sum_i \mathbb{E}(X_i^2) + \sum_{i \neq j} \mathbb{E}(X_i X_j)$$

and

$$(\mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n))^2 = \mathbb{E}(X_1)^2 + \mathbb{E}(X_2)^2 + \dots + \mathbb{E}(X_n)^2 + \sum_{i \neq j} \mathbb{E}(X_i) \mathbb{E}(X_j)$$

Since

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \mathbb{E}((X_1 + X_2 + \dots + X_n)^2) - (\mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n))^2$$

the formula follows

**Theorem 2.2.1** *If  $X, Y$  are independent stochastic variables and  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  are measurable with respect to  $\mathcal{B}$  then the composite functions  $f(X)$  and  $g(Y)$  are independent random variables and  $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$*

Since both  $f$  and  $X$  are measurable the composite map  $f(X)$  is also measurable (*Exercise: Prove this*). Thus  $f(X)$  is a random variable. Let  $A, B \subset \mathbb{R}$  be Borel sets then  $f^{-1}(A)$  and  $g^{-1}(B)$  are Borel sets and  $(f(X))^{-1}(A) = X^{-1}(f^{-1}(A))$ ,  $(g(Y))^{-1}(B) = Y^{-1}(g^{-1}(B))$  are independent events. Since  $\sigma(f(X))$  (resp.  $\sigma(g(Y))$ ) is generated by sets of the form  $(f(X))^{-1}(A)$  (resp.  $(g(Y))^{-1}(B)$ ) where  $A$  (resp.  $B$ ) runs through the Borel sets, it follows that  $f(X)$  and  $g(Y)$  are independent.

To prove the second part assume first that  $f$  and  $g$  are simple functions, say  $f$  takes the value  $y_k$  on the Borel set  $B_k$ ,  $k = 1, 2, 3, \dots$  and  $g$  takes the value  $z_j$  on the Borel set  $C_j$ ,  $j = 1, 2, 3, \dots$ . Then both  $f(X)$ ,  $g(Y)$  and  $f(X)g(Y)$  are simple functions. We have  $\int_{\Omega} f(X) d\mu = \sum_k y_k \mu(X^{-1}(B_k))$ ,  $\int_{\Omega} g(Y) d\mu = \sum_j z_j \mu(Y^{-1}(C_j))$  and  $\int_{\Omega} f(X)g(Y) d\mu = \sum_k \sum_j y_k z_j \mu(X^{-1}(B_k) \cap Y^{-1}(C_j))$ .

Since  $X$  and  $Y$  are independent, the events  $X^{-1}(B_k)$  and  $Y^{-1}(C_j)$  are independent so  $\mu(X^{-1}(B_k) \cap Y^{-1}(C_j)) = \mu(X^{-1}(B_k))\mu(Y^{-1}(C_j))$  and hence

the double sum is equal to  $\sum_k y_k \mu(X^{-1}(B_k)) \sum_j z_j \mu(Y^{-1}(C_j))$ . This proves the statement when  $f$  and  $g$  are simple functions.

To prove it in general choose sequences of simple functions  $\{f_n\}$  and  $\{g_n\}$  converging uniformly to  $f$  and  $g$ . Then  $f_n(X) \rightarrow f(X)$ ,  $g_n(Y) \rightarrow g(Y)$  and  $f_n(X)g_n(Y) \rightarrow f(X)g(Y)$ , uniformly (*Exercise: Prove this*). But by definition of the Lebesgue integral  $\int_{\Omega} f_n(X) d\mu \rightarrow \int_{\Omega} f(X) d\mu$  etc.

**Corollary 2.2.1** *If  $X, Y$  are independent then  $Cov(X, Y) = 0$*

By taking  $f = g = id$  in theorem 2.2.1 we get  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$  hence  $Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$

**Corollary 2.2.2** *If  $X_1, X_2, \dots, X_n$  are pairwise independent then  $Var(X_1 + X_2 + \dots + X_n) = Var(X_1) + Var(X_2) + \dots + Var(X_n)$*

*Warning: It is not true that  $Cov(X, Y) = 0$  implies that  $X, Y$  are independent*

## 2.3 Distributions and Density functions

Let now  $(\Omega, \mathcal{F}, \mu)$  be a probability space and  $X$  a random variable on  $\Omega$ . Since  $X$  is, by definition, a measurable function we know that for all  $c \in \mathbb{R}$ , the set  $\{\omega | X(\omega) < c\}$  is measurable, it is customary to simply write  $\{X < c\}$  for this set.

**Definition 2.3.1** *The distribution of the random variable  $X$  is the function  $D_X(c) = \mu(\{X < c\})$ .*

Thus  $D_X : \mathbb{R} \rightarrow [0, 1]$ .

**Example 2.3** *Let  $\alpha$  be a positive real number and consider the function on the positive integers  $n \mapsto e^{-\alpha} \alpha^n / n!$ . A random variable  $X$  taking positive integer values (thus  $X$  is a simple function) is Poisson distributed with parameter  $\alpha$  if  $\mu(X = n) = e^{-\alpha} \alpha^n / n!$ . The mean of a Poisson distributed random variable is  $\mathbb{E}(X) = \sum_{n=1} n e^{-\alpha} \alpha^n / n! = e^{-\alpha} \alpha \sum_{n=1} \alpha^{n-1} / (n-1)! = \alpha e^{-\alpha} e^{\alpha} = \alpha$ .*

The variance is given by  $\mathbb{E}(X^2) - \alpha^2$ . We have

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{n=1}^{\infty} n^2 \mu(X^2 = n^2) \\
&= \sum_{n=1}^{\infty} n^2 \mu(X = n) \\
&= \sum_{n=1}^{\infty} n^2 e^{-\alpha} \alpha^n / n! \\
&= e^{-\alpha} \alpha \sum_{n=1}^{\infty} n \alpha^{n-1} / (n-1)! \\
&= e^{-\alpha} \alpha \frac{d}{d\alpha} \sum_{n=1}^{\infty} \alpha^n / (n-1)! \\
&= e^{-\alpha} \alpha \frac{d}{d\alpha} \alpha e^{\alpha} \\
&= \alpha + \alpha^2
\end{aligned}$$

Thus  $\text{Var}(X) = \alpha$

Let  $X : \Omega \rightarrow \mathbb{R}$  be a random variable. For every Borel set  $B$ ,  $X^{-1}(B) \in \mathcal{F}$  and hence  $\mu(X^{-1}(B))$  is defined.

**Lemma 2.3.1** *The set function  $\mu_X : \mathcal{B} \rightarrow \mathbb{R}$ ,  $B \mapsto \mu(X^{-1}(B))$  defines a measure on the Borel sets in  $\mathbb{R}$*

Let  $x$  denote the Lebesgue measure on  $\mathcal{B}$  and assume that the measure  $\mu_X$  is absolutely continuous with respect to  $x$  i.e. if  $A$  is a Borel set of measure 0 then  $X^{-1}(A)$  is a subset of measure 0 in  $\mathcal{F}$ . By the Radon-Nikodym theorem there exists an integrable function  $\phi_X : \mathbb{R} \rightarrow \mathbb{R}$  such that for all Borel sets  $B$  we have  $\mu_X(B) = \int_B \phi_X(x) dx$ . (*Remark:  $\mathbb{R}$  does not have finite Lebesgue measure. However under the measure  $\mu_X$ ,  $\mathbb{R}$  has measure 1 and this suffices to be able to apply the Radon-Nikodym theorem.*)

**Definition 2.3.2** *The function  $\phi_X$  is called the density of the stochastic variable. It is of course only defined for stochastic variables  $X$  for which the measure  $\mu_X$  is absolutely continuous with respect to the Lebesgue measure.*

**Remark 2.1** *If  $X$  has density  $\phi_X$  then  $D_X(c) = \int_{-\infty}^c \phi_X(x) dx$  and  $\int_{\mathbb{R}} \phi_X(x) dx = \mu(X^{-1}(\mathbb{R})) = \mu(\Omega) = 1$ . If  $\Omega$  is a finite probability space*



there are no stochastic variables with density function because in this case the measure  $\mu_X$  is not absolutely continuous with respect to the Lebesgue measure (Exercise: Show this)

**Example 2.4** The most famous density function is the Gaussian with mean  $\mu$  and variance  $\sigma^2$ :  $\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$ . The associated distribution is the Normal Distribution (with mean  $\mu$  and variance  $\sigma^2$ ). If  $X$  is a random variable with this distribution we write  $X \sim N(\mu, \sigma^2)$ . If  $\mu = 0$  and  $\sigma = 1$  then the distribution is called the Standard Normal Distribution and a random variable with this distribution is said to be standard normal.

**Theorem 2.3.1** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be measurable with respect to the  $\sigma$ -algebra  $\mathcal{B}$  of Borel sets. Let  $X$  be a random variable with density function  $\phi_X$ . Then  $\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x)\phi_X(x)dx$

Assume first that  $f$  is a simple function, say  $f(x) = y_k$  on the Borel set  $B_k$ ,  $k = 1, 2, 3, \dots$ . Then  $f(X)$  is also a simple function taking the value  $y_k$  on the measurable set  $X^{-1}(B_k)$ . Thus  $\int_{\Omega} f(X) d\mu = \sum_k y_k \mu(X^{-1}(B_k)) = \sum_k y_k \int_{B_k} \phi_X(x)dx = \sum_k \int_{B_k} f(x)\phi_X(x)dx = \int_{\mathbb{R}} f(x)\phi_X(x)dx$ . This proves the theorem when  $f$  is a simple function.

To prove the theorem in general, let  $\{f_n\}$  be a sequence of simple functions converging uniformly to  $f$ . Then  $\{f_n(X)\}$  is a sequence of simple functions (on  $\Omega$ ) converging uniformly to  $f(X)$  (Exercise: Prove this). But then  $\mathbb{E}(f_n(X)) \rightarrow \mathbb{E}(f(X))$  and  $\int_{\mathbb{R}} f_n(x)\phi_X(x)dx \rightarrow \int_{\mathbb{R}} f(x)\phi_X(x)dx$ .

In particular we get by taking  $f = id$  that  $\mathbb{E}(X) = \int_{\mathbb{R}} x \phi_X(x)dx$  and by taking  $f(x) = x^n$  we get  $\mathbb{E}(X^n) = \int_{\mathbb{R}} x^n \phi_X(x)dx$ . The number  $\mathbb{E}(X^n)$  is called the  $n$ 'th moment of  $X$ .

**Definition 2.3.3** The Moment Generating Function or The Characteristic Function of the random variable  $X$  is the power series  $f_X(s) = \sum_{n=0}^{\infty} \frac{(is)^n}{n!} \mathbb{E}(X^n) = \mathbb{E}(\exp(is \cdot X))$ . Remark that  $f_X$  is a complex valued function.

If  $X$  has density function  $\phi_X$  then we get  $f_X(s) = \int_{\mathbb{R}} \exp(is \cdot x)\phi_X(x)dx$ . This is the Fourier transform of the density function. Using the inverse Fourier transform we get  $\phi_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} f_X(s)\exp(-ix \cdot s)ds$ . Thus the characteristic function and the density function determine each other.

**Example 2.5** Let  $X \sim N(0, 1)$ . Then we have

$$\begin{aligned} f_X(s) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(is \cdot x) \exp(-x^2/2) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-(x - is)^2/2) \exp(-s^2/2) dx = \exp(-s^2/2) \end{aligned}$$

Expanding  $\exp(-s^2/2)$  in a powerseries and comparing coefficients we get the following formula for the moments of  $X$ :

$$\mathbb{E}(X^{2n+1}) = 0 \tag{1}$$

$$\mathbb{E}(X^{2n}) = \frac{(2n)!}{2^n n!} \tag{2}$$

**Example 2.6** Let  $X \sim N(0, 1)$ . Let  $Z$  be independent of  $X$  with distribution  $\mu(\{Z = 1\}) = \mu(\{Z = -1\}) = 1/2$ . Consider  $Y = ZX$  and compute  $\text{Cov}(X, Y) = \mathbb{E}(ZX^2) - \mathbb{E}(X)\mathbb{E}(ZX) = \mathbb{E}(ZX^2)$  because  $\mathbb{E}(X) = 0$ . Since  $X$  and  $Z$  are independent  $X^2$  and  $Z$  are also independent and  $\mathbb{E}(ZX^2) = \mathbb{E}(Z)\mathbb{E}(X^2)$ . Now  $\mathbb{E}(Z) = \int_{\Omega} Z d\mu = \int_{\{Z=-1\}} -1 d\mu + \int_{\{Z=1\}} 1 d\mu = -1/2 + 1/2 = 0$ . Thus  $\text{Cov}(X, Y) = 0$ , but if they were independent then  $X^2$  and  $Y^2$  would also be independent, however they are equal almost everywhere and so they are certainly not independent. This example shows that vanishing of covariance does not imply independence.

Consider two random variables  $X, Y : \Omega \longrightarrow \mathbb{R}$ . Then  $(X, Y)$  induce a measure on  $\mathcal{B}(\mathbb{R}^2)$  (strictly speaking we have not defined the  $\sigma$ -algebra of Borel sets in  $\mathbb{R}^2$  but it is clear what it should be: the  $\sigma$ -algebra generated by sets of the form  $\{(x, y) \in \mathbb{R}^2 \mid x < a, y < b\}$ ) by  $\mu_{X,Y}(C) = \mu(\{(X, Y) \in C\})$ , for all  $C \in \mathcal{B}(\mathbb{R}^2)$ . The *joint density* is a function  $\phi_{X,Y} : \mathbb{R}^2 \longrightarrow \mathbb{R}$  such that  $\mu_{X,Y}(C) = \int_C \phi_{X,Y}(x, y) dx dy$  for all  $C \in \mathcal{B}(\mathbb{R}^2)$ . Even if both  $X$  and  $Y$  have densities  $(X, Y)$  may not have a joint density.

Let  $k : \mathbb{R}^2 \longrightarrow \mathbb{R}$  be measurable with respect to Borel sets. Then  $k(X, Y) : \Omega \longrightarrow \mathbb{R}$  is a random variable and we have

$$\mathbb{E}(k(X, Y)) = \int_{\mathbb{R}^2} k(x, y) \phi_{X,Y}(x, y) dx dy$$

This is proved exactly as in the one-variable case.

If  $X, Y$  have joint density  $\phi_{X,Y}$  both  $X$  and  $Y$  have a density. In fact we have

$$\begin{aligned}\mu_Y(B) &= \mu(\{Y \in B\}) \\ &= \mu(\{(X, Y) \in \mathbb{R} \times B\}) \\ &= \int_{\mathbb{R} \times B} \phi_{X,Y} dx dy \\ &= \int_B \int_{\mathbb{R}} \phi_{X,Y}(x, y) dx dy\end{aligned}$$

Thus  $\phi_Y(y) = \int_{\mathbb{R}} \phi(x, y) dx$  is the density of  $Y$  (in this case it is sometimes called the marginal density). *Exercise: Formulate the notion of joint density and marginal density for any number of random variables  $X_1, X_2, \dots, X_n : \Omega \rightarrow \mathbb{R}$*

**Proposition 2.3.1** *Assume  $X, Y$  are independent random variables with densities  $\phi_X$  and  $\phi_Y$ . Then  $X, Y$  have joint density  $\phi_{X,Y}(x, y) = \phi_X(x)\phi_Y(y)$*

Since  $\mathcal{B}(\mathbb{R}^2)$  is generated by sets of the form  $A \times B$  with  $A, B \in \mathcal{B}(\mathbb{R})$  it suffices to check that  $\mu_{X,Y}(A \times B) = \int_{A \times B} \phi_X(x)\phi_Y(y) dx dy$ . But we have  $\mu_{X,Y}(A \times B) = \mu(\{X \in A\} \cap \{Y \in B\}) = \mu(\{X \in A\})\mu(\{Y \in B\}) = \mu_X(A)\mu_Y(B)$  because the  $\sigma$ -algebras generated by  $X$  and  $Y$  are independent. On the other hand  $\int_{A \times B} \phi_X(x)\phi_Y(y) dx dy = \int_A \int_B \phi_X(x)\phi_Y(y) dx dy = \int_A \phi_X(x) dx \int_B \phi_Y(y) dy = \mu_X(A)\mu_Y(B)$  *Exercise: Generalize this result to  $n$  independent random variables  $X_1, X_2, \dots, X_n$  with density functions  $\phi_{X_1}, \phi_{X_2}, \dots, \phi_{X_n}$*

**Definition 2.3.4** *Let  $\{X_n\}$  be a sequence of random variables on a probability space  $(\Omega, \mathcal{F}, \mu)$  and let  $X$  be a random variable on the same space. We say that  $X_n \rightarrow X$  in probability if for all  $\varepsilon > 0$ ,  $\mu(|X_n - X| > \varepsilon) \rightarrow 0$ .*

**Theorem 2.3.2 (Law of Large Numbers)** *Let  $X_1, X_2, \dots$  be a sequence of independent, identically distributed random variables, each with  $\mathbb{E}(X_i) = m$  and  $\text{Var}(X_i) = \sigma^2$ . Define  $Y_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ . Then  $Y_n \rightarrow m$  in probability.*

By Chebychef's inequality we have

$$\mu(|Y_n - m| > \varepsilon) = \mu(|Y_n - m|^2 > \varepsilon^2) \leq \frac{1}{\varepsilon^2} \int_{\Omega} |Y_n - m|^2 d\mu$$

Thus we only need to prove  $\int_{\Omega} |Y_n - m|^2 d\mu \rightarrow 0$ . But  $\int_{\Omega} |Y_n - m|^2 d\mu = \mathbb{E}((Y_n - m)^2) = \text{Var}(Y_n)$  and since the  $X_i$ 's are independent  $\text{Var}(Y_n) = \frac{1}{n^2}(\text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_n)) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \rightarrow 0$

**Theorem 2.3.3 The Central Limit Theorem** *Let  $X_1, X_2, \dots$  be a sequence of independent, identically distributed random variables, each with  $\mathbb{E}(X_i) = \mu$  and  $\text{Var}(X_i) = \sigma^2$ . Set  $Z_n = \frac{(X_1 - \mu) + (X_2 - \mu) + \cdots + (X_n - \mu)}{\sqrt{n}}$ . Then  $\mathbb{E}(Z_n) = 0$  and  $\text{Var}(Z_n) = \sigma^2$ . Let  $A \in \mathcal{B}(\mathbb{R})$  then  $\lim_{n \rightarrow \infty} \mu(Z_n \in A) = \frac{1}{\sigma\sqrt{2\pi}} \int_A \exp(-\frac{x^2}{2\sigma^2}) dx$ . In other words the distribution of  $Z_n$  tends to the normal distribution with mean 0 and variance  $\sigma^2$*

The Central Limit Theorem comes in different flavors. The assumptions we have stated here can be weakened considerably. For instance it is not necessary to assume that the random variables have identical distributions or have the same variance (but they do have to have the same expected value). Let for simplicity  $\mathbb{E}(X_i) = 0$  and let  $\text{Var}(X_i) = b_i^2$ . Assume that  $X_i$  has density  $p_i$ . Let  $\sigma_n^2 = \text{Var}(Z_n) = \frac{1}{n} \sum b_i^2$ . We further assume that the following condition (known as the *Lindeberg condition*) is satisfied.

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_n^2} \sum_{i=1}^n \int_{|x| > t\sigma_n} x^2 p_i(x) dx = 0$$

for any fixed  $t > 0$ .

Under these conditions the distribution of  $Z_n/\sigma_n$  converges to a standard normal distribution.

In practice these conditions are rather weak and are satisfied for most density functions. An exception is the *Cauchy* distribution  $p_i(x) = a_i \frac{1}{\pi(x^2 + a_i^2)}$ . Here the Lindeberg condition is not satisfied (*Exercise: Verify this*)

## 2.4 Conditional Expectation

Let  $(\Omega, \mathcal{F}, \mu)$  be a probability space and  $X$  a random variable such that  $\mathbb{E}(X)$  exists. Let  $\mathcal{G} \subset \mathcal{F}$  be a sub- $\sigma$ -algebra. While  $X$  certainly is measurable with respect to  $\mathcal{F}$  it may not be measurable with respect to  $\mathcal{G}$ . The conditional expectation  $\mathbb{E}(X|\mathcal{G})$  is a random variable with respect to  $\mathcal{G}$  such that for all  $A \in \mathcal{G}$  we have  $\int_A \mathbb{E}(X|\mathcal{G}) d\mu = \int_A X d\mu$

**Example 2.7** If  $\mathcal{G}$  is the trivial  $\sigma$ -algebra,  $\mathcal{G} = \{\emptyset, \Omega\}$ , the only functions which are random variables with respect to  $\mathcal{G}$  are constant. Since the only non-empty set in  $\mathcal{G}$  is  $\Omega$  we get  $\int_{\Omega} X d\mu = \mathbb{E}(X) = \int_{\Omega} \mathbb{E}(X) d\mu$  hence  $\mathbb{E}(X|\mathcal{G})$  is the constant function  $\mathbb{E}(X)$

Before we prove existence of the conditional expectation let us revisit the binomial tree of stock prices. Recall that the stock price at time  $t = i$ ,  $S_i$ , is a random variable with respect to the  $\sigma$ -algebra,  $\mathcal{F}_i$  generated by the sets of the form  $U_{\eta} = \{\omega \in \Omega_n | (\omega_1, \omega_2, \dots, \omega_i) = \eta\}$ , where  $\eta \in \Omega_i$ . Clearly we have  $\mathcal{F}_i \subset \mathcal{F}_j$  if  $i \leq j$ . We assume that the probability of an up move is  $p$  and the probability for a down move is  $q = 1 - p$ . In other words the probability measure on  $\Omega_n$  is defined by the probability of 1 being  $p$  and the probability of 0 being  $q$ . Let  $\eta \in \Omega_i$  and assume  $\eta$  consists of  $k$  1's and  $i - k$ , 0's. We can write  $U_{\eta} = U_{\eta,0} \cup U_{\eta,1}$ . On  $U_{\eta,0}$ ,  $S_{i+1}$  is constant with value  $u^k d^{i+1-k} S_0$  and on  $U_{\eta,1}$  it is constant with value  $u^{k+1} d^{i-k} S_0$ . We have  $\mu(U_{\eta,0}) = p^k q^{i+1-k}$  and  $\mu(U_{\eta,1}) = p^{k+1} q^{i-k}$ . It follows that  $\int_{U_{\eta}} S_{i+1} d\mu = u^k d^{i+1-k} S_0 p^k q^{i+1-k} + u^{k+1} d^{i-k} S_0 p^{k+1} q^{i-k} = (pu + qd)(u^k d^{i-k} S_0 p^k q^{i-k})$ . Now  $S_i$  is constant on  $U_{\eta}$  with value  $u^k d^{i-k} S_0$  and  $\mu(U_{\eta}) = p^k q^{i-k}$  hence  $\int_{U_{\eta}} S_i = u^k d^{i-k} p^k q^{i-k} S_0$ . It follows that  $\int_{U_{\eta}} S_{i+1} d\mu = \int_{U_{\eta}} (pu + qd) S_i d\mu$ . Since  $\mathcal{F}_i$  is generated by the sets  $U_{\eta}$  as  $\eta$  runs through  $\Omega_i$  we get  $\int_A S_{i+1} d\mu = \int_A (pu + qd) S_i d\mu$  for all  $A \in \mathcal{F}_i$ . This shows that  $\mathbb{E}(S_{i+1} | \mathcal{F}_i) = (pu + qd) S_i$ . In general we get for  $j > i$ ,  $\mathbb{E}(S_j | \mathcal{F}_i) = (pu + qd)^{j-i} S_i$  (*Exercise: Prove this*)

In case  $\Omega$  is a finite set we can find an explicit formula for the conditional expectation. Assume that  $\Omega$  is a finite set,  $\mathcal{F}$  a  $\sigma$ -algebra and  $\mu$  a probability measure. For every  $\omega \in \Omega$  we consider  $f(\omega) = \bigcap_{\omega \in A \in \mathcal{F}} A$ . Clearly  $f(\omega)$  is the smallest set in  $\mathcal{F}$  containing  $\omega$ . Remark that this construction works because  $\Omega$  is finite and hence we are only dealing with an intersection of finitely many sets.

**Lemma 2.4.1** *The sets  $\{f(\omega)\}_{\omega \in \Omega}$  form a partition of  $\Omega$  i.e.  $f(\omega) \neq f(\omega')$  implies  $f(\omega) \cap f(\omega') = \emptyset$  and  $\Omega = \bigcup f(\omega)$*

Since  $\omega \in f(\omega)$  it is clear that the union is all of  $\Omega$ . Assume  $f(\omega) \neq f(\omega')$ . If  $\omega \in f(\omega')$  we have  $f(\omega) \subset f(\omega')$  because by construction  $f(\omega)$  is contained in every set in  $\mathcal{F}$  which contains  $\omega$ . Similarly if  $\omega' \in f(\omega)$  we have  $f(\omega') \subset f(\omega)$ . Since the two sets are different we must have either  $\omega \notin f(\omega')$  or  $\omega' \notin f(\omega)$ . In the first case we have  $\omega \in f(\omega) - f(\omega') \subset f(\omega)$  and since the difference set

is in  $\mathcal{F}$  we must have  $f(\omega) - f(\omega') = f(\omega)$  which implies  $f(\omega) \cap f(\omega') = \emptyset$ . The other case is treated exactly the same.

Thus a  $\sigma$ -algebra gives rise to a partition. On the other hand given a partition we can construct a  $\sigma$ -algebra by taking all possible unions of sets in the partition (*Exercise: Verify that this gives a  $\sigma$ -algebra and that the partition associated to this  $\sigma$ -algebra is the partition we started with*)

**Lemma 2.4.2** *Let  $X$  be a function on  $\Omega$  then  $X$  is a random variable with respect to  $\mathcal{F}$  if and only if  $X$  is constant on the sets  $f(\omega)$*

Indeed assume  $X$  is measurable with respect to  $\mathcal{F}$ . If  $X(\omega) = a$  then  $\omega \in f(\omega) - \{X \neq a\}$ . This set is in  $\mathcal{F}$  hence  $f(\omega) \subset f(\omega) - \{X \neq a\}$  so  $f(\omega) \cap \{X \neq a\} = \emptyset$  which shows that  $X = a$  on  $f(\omega)$ . Conversely if  $X$  is constant on the  $f(\omega)$ 's, then  $\{X = a\} = \bigcup_{\{\omega | X(\omega) = a\}} f(\omega) \in \mathcal{F}$ . Hence  $X$  is measurable with respect to  $\mathcal{F}$

**Lemma 2.4.3** *Let  $\mathcal{G}$  and  $\mathcal{F}$  be  $\sigma$ -algebras. Then  $\mathcal{G} \subset \mathcal{F}$  if and only if  $f(\omega) \subset g(\omega)$  for all  $\omega \in \Omega$*

If  $\mathcal{G} \subset \mathcal{F}$  then  $A \in \mathcal{G}$  implies  $A \in \mathcal{F}$ , hence  $f(\omega) = \bigcap_{\omega \in A \in \mathcal{F}} A \subset \bigcap_{\omega \in A \in \mathcal{G}} A = g(\omega)$ . To go the other way notice that if  $\omega' \in g(\omega)$  then  $g(\omega') = g(\omega)$  and so  $g(\omega) = \bigcup_{\omega' \in g(\omega)} f(\omega') \in \mathcal{F}$

**Theorem 2.4.1** *Let  $\Omega$  be a finite probability space and let  $\mathcal{G} \subset \mathcal{F}$ . Let  $X$  be a random variable with respect to  $\mathcal{F}$ . Then*

$$\mathbb{E}(X|\mathcal{G})(\omega) = \frac{\int_{g(\omega)} X d\mu}{\mu(g(\omega))} = \frac{\sum_{\omega' \in g(\omega)} X(\omega')\mu(\omega')}{\sum_{\omega' \in g(\omega)} \mu(\omega')}$$

By definition of  $\mathbb{E}(X|\mathcal{G})$  we have  $\int_{g(\omega)} X d\mu = \int_{g(\omega)} \mathbb{E}(X|\mathcal{G}) d\mu$ . Since  $\mathbb{E}(X|\mathcal{G})$  is measurable with respect to  $\mathcal{G}$  it is constant on  $g(\omega)$  hence  $\int_{g(\omega)} \mathbb{E}(X|\mathcal{G}) d\mu = \mathbb{E}(X|\mathcal{G})(\omega)\mu(g(\omega)) = \mathbb{E}(X|\mathcal{G})(\omega) \sum_{\omega' \in g(\omega)} \mu(\omega')$ .

On the other hand  $\int_{g(\omega)} X d\mu = \sum_{\omega' \in g(\omega)} X(\omega')\mu(\omega')$ . This proves the theorem

**Example 2.8** Let  $X, Y$  be random variables (with respect to  $\mathcal{F}$ ). Then the conditional expectation  $\mathbb{E}(X|Y)$  is defined to be  $\mathbb{E}(X|\sigma(Y))$ . Assume  $X, Y$  have joint density  $\phi_{X,Y}$ . We define the conditional density  $\phi_{X|Y}(x|y) = \frac{\phi_{X,Y}(x,y)}{\phi_Y(y)}$ . Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be measurable and define  $g : \mathbb{R} \rightarrow \mathbb{R}$  by  $g(y) = \int_{\mathbb{R}} h(x)\phi_{X|Y}(x|y)dx$ . Then we have  $\mathbb{E}(h(X)|Y)(\omega) = g(Y(\omega))$ .

To see this let  $A = \{Y \in B\}$ ,  $B \in \mathcal{B}(\mathbb{R})$ , be a set in  $\sigma(Y)$  and let  $\chi_A$  and  $\chi_B$  denote the characteristic functions of these sets (the characteristic function of a set is the function which is constant with value 1 on the set and constant with value 0 on the complement of the set). Then we have

$$\begin{aligned}
\int_A \mathbb{E}(h(X)|Y)d\mu &= \int_A h(X)d\mu \\
&= \int_{\Omega} \chi_A h(X)d\mu \\
&= \int_{\Omega} \chi_B(Y)h(X)d\mu \\
&= \mathbb{E}(h(X)\chi_B(Y)) \\
&= \int_{\mathbb{R}^2} h(x)\chi_B(y)\phi_{X,Y}(x,y)dxdy \\
&= \int_{\mathbb{R}} \chi_B(y) \int_{\mathbb{R}} h(x)\phi_{X,Y}(x,y)dxdy \\
&= \int_B \int_{\mathbb{R}} h(x)\phi_{X,Y}(x,y)dxdy \\
&= \int_B \int_{\mathbb{R}} h(x)\phi_{X|Y}(x|y)\phi_Y(y)dxdy \\
&= \int_B g(y)\phi_Y(y)dy \\
&= \int_A g(Y)d\mu
\end{aligned}$$

Since  $g(Y)$  is measurable with respect to  $\sigma(Y)$  this shows that  $\mathbb{E}(h(X)|Y) = g(Y)$ . The function  $g$  is sometimes written  $g(y) = \mathbb{E}(h(X)|Y = y)$

**Theorem 2.4.2** Let  $\mathcal{G} \subset \mathcal{F}$  and let  $Y$  be measurable with respect to  $\mathcal{G}$  and  $X$  measurable with respect  $\mathcal{F}$ . Then  $\mathbb{E}(YX|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$

This theorem is proved in the usual way. Assume first that  $Y$  is a simple function  $Y = \sum c_i \chi_{A_i}$ ,  $A_i \in \mathcal{G}$ ,  $i = 1, 2, \dots$ . Then for all  $B \in \mathcal{G}$ ,  $\int_B Y X d\mu = \sum c_i \int_B \chi_{A_i} X = \sum c_i \int_{A_i \cap B} X = \sum c_i \int_{A_i \cap B} \mathbb{E}(X|\mathcal{G}) = \int_B Y \mathbb{E}(X|\mathcal{G})$ . This proves the result in the case  $Y$  is a simple function.

For a general  $Y$  we can find a sequence of simple functions  $\{Y_n\}$  converging uniformly to  $Y$  and we have  $\mathbb{E}(YX|\mathcal{G}) = \lim \mathbb{E}(Y_n X|\mathcal{G}) = \lim Y_n \mathbb{E}(X|\mathcal{G}) = Y \mathbb{E}(X|\mathcal{G})$

The proof of the existence of the conditional expectation is not very illuminating but we include it for the sake of completeness.

Assume first that  $X$  is non-negative i.e.  $X(\omega) \geq 0$  for all  $\omega \in \Omega$ . Then  $\int_A X d\mu \geq 0$  for all  $A \in \mathcal{G}$  and so  $A \mapsto \nu(A) = \int_A X d\mu$  defines a measure on  $\mathcal{G}$ . By the Radon-Nikodym theorem there exists a function  $Y = \frac{d\nu}{d\mu}$ , measurable with respect to  $\mathcal{G}$ , such that  $\nu(A) = \int_A Y d\mu$  for all  $A \in \mathcal{G}$ . This proves the existence in the case where  $X$  is non-negative.

In the general case we use the trick from Theorem 1.5.4 (The change of measure formula) and consider  $X^+ = \max(X, 0)$  and  $X^- = \max(-X, 0)$  then both  $X^+$  and  $X^-$  are non-negative and  $X = X^+ - X^-$ . By the previous case, there exists  $Y^+$  and  $Y^-$  as above and then  $Y = Y^+ - Y^-$  will work (*Exercise: Fill in the details of this part of the proof*). This way of writing a function as a difference of two non-negative functions is standard in integration theory and often makes it possible to apply the Radon-Nikodym theorem in the same way as here.

**Theorem 2.4.3 (Transitivity)** *Let  $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$  and let  $X$  be a random variable with respect to  $\mathcal{F}$ . Then  $\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})$  a.e.*

By definition  $\int_A \mathbb{E}(X|\mathcal{H}) d\mu = \int_A X d\mu$  for all  $A \in \mathcal{H}$ . Also  $\int_A X d\mu = \int_A \mathbb{E}(X|\mathcal{G}) d\mu$  for all  $A \in \mathcal{G}$  and  $\int_A \mathbb{E}(X|\mathcal{G}) d\mu = \int_A \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) d\mu$  for all  $A \in \mathcal{H}$ . This shows that  $\int_A \mathbb{E}(X|\mathcal{H}) d\mu = \int_A \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) d\mu$  for all  $A \in \mathcal{H}$  and hence  $\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})$  a.e.

### 3 Stochastic Processes

Let  $(\Omega, \mathcal{F}, \mu)$  be a probability space (in finance this is often called the *state space*).



Basically a stochastic process is an indexed set of random variables  $\{X_t\}_{t \in I}$ . The index set  $I$  is either a finite set or an interval in  $\mathbb{R}$  and is usually thought of as the time variable.

Assume that for any finite set of indices  $\{t_1, t_2, \dots, t_n\} \subset I$ , the joint density  $\phi_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}$  exists. For an ordered set of indices  $t_1 \geq t_2 \geq \dots \geq t_k \geq t_{k+1} \geq t_{k+2} \geq \dots \geq t_n$  we define the conditional density

$$\phi_{X_{t_1}, \dots, X_{t_k} | X_{t_{k+1}}, \dots, X_{t_n}} = \phi_{X_{t_1}, X_{t_2}, \dots, X_{t_n}} / \phi_{X_{t_{k+1}}, X_{t_{k+2}}, \dots, X_{t_n}}$$

The simplest kind of stochastic process is that of complete independence i.e.  $\phi_{X_{t_1}, X_{t_2}, \dots, X_{t_n}} = \prod \phi_{X_{t_i}}$ , which means that the value of  $X_t$  is completely independent of its values in the past. An even more special case occurs when  $\phi_{X_t}$  is independent of  $t$ . We then have a situation in which the same experiment is repeated at successive times.

The next simplest kind of stochastic process is the *Markov process*. It is characterized by the fact that the conditional probability only depends on the most recent condition i.e. if  $t_1 \geq t_2 \geq \dots \geq t_n \geq \tau_1 \geq \tau_2 \geq \dots \geq \tau_m$  then  $\phi_{X_{t_1}, X_{t_2}, \dots, X_{t_n} | X_{\tau_1}, X_{\tau_2}, \dots, X_{\tau_m}} = \phi_{X_{t_1}, X_{t_2}, \dots, X_{t_n} | X_{\tau_1}}$

We have the identity

$$\begin{aligned} \phi_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) &= \phi_{X_{t_1} | X_{t_2}, \dots, X_{t_n}}(x_1 | x_2, \dots, x_n) \phi_{X_{t_2}, \dots, X_{t_n}}(x_2, \dots, x_n) \\ &= \phi_{X_{t_1} | X_{t_2}}(x_1 | x_2) \phi_{X_{t_2}, X_{t_3}, \dots, X_{t_n}}(x_2, x_3, \dots, x_n) \end{aligned}$$

where the second equality follows from the Markov property. By induction we get for  $t_1 \geq t_2 \geq \dots \geq t_n$

$$\phi_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) = \phi_{X_{t_1} | X_{t_2}}(x_1 | x_2) \dots \phi_{X_{t_{n-1}} | X_{t_n}}(x_{n-1} | x_n) \phi_{X_{t_n}}(x_n)$$

When we discussed marginal density functions we saw that we can write  $\phi_{X_1, X_3}(x, z) = \int_{\mathbb{R}} \phi_{X_1, X_2, X_3}(x, y, z) dy$ . Hence we can write

$$\begin{aligned} \phi_{X_1 | X_3}(x | z) &= \phi_{X_1, X_3}(x, z) / \phi_{X_3}(z) \\ &= \int_{\mathbb{R}} \phi_{X_1, X_2, X_3}(x, y, z) / \phi_{X_3}(z) dy \\ &= \int_{\mathbb{R}} \phi_{X_1, X_2, X_3}(x, y, z) / \phi_{X_2, X_3}(y, z) \phi_{X_2, X_3}(y, z) / \phi_{X_3}(z) dy \\ &= \int_{\mathbb{R}} \phi_{X_1 | X_2, X_3}(x | y, z) \phi_{X_2 | X_3}(y | z) dy \end{aligned}$$

Applying this formula to a Markov process we get for  $t_1 \leq t_2 \leq t_3$ :

$$\begin{aligned}\phi_{X_{t_1}|X_{t_3}}(x|z) &= \int_{\mathbb{R}} \phi_{X_{t_1}|X_{t_2},X_{t_3}}(x|y,z)\phi_{X_{t_2}|X_{t_3}}(y|z)dy \\ &= \int_{\mathbb{R}} \phi_{X_{t_1}|X_{t_2}}(x|y)\phi_{X_{t_2}|X_{t_3}}(y|z)dy\end{aligned}$$

where the last equality follows from the Markov property. This equation is known as the *Chapman-Kolmogorov equation*.

**Example 3.1** *The most important stochastic process is the Wiener Process also known as Brownian Motion,  $\{W_t\}_{t \in I}$  ( $I = [0, T]$ ) or  $I = \{0, 1, \dots, N\}$ ). It is the Markov process given by the conditional densities*

$$\phi_{W_{t_1}|W_{t_2}}(x_1|x_2) = \frac{1}{\sqrt{2\pi D(t_1 - t_2)}} \exp\left(\frac{-(x_1 - x_2)^2}{2D(t_1 - t_2)}\right)$$

and the initial condition  $W_0 = 0$ . We can compute the density of the stochastic variable  $W_t$  as follows: we have  $\phi_{W_t}(x) = \int_{-\infty}^{\infty} \phi_{W_t, W_0}(x, y) dy$  and  $\phi_{W_t, W_0} = \phi_{W_t|W_0}\phi_{W_0}$ . There is of course the problem that the density  $\phi_{W_0}$  does not exist. It would have to be a function such that  $\int_{-\infty}^{\infty} \phi_{W_0}(v) dv = 1$  but the integral over any interval not containing 0 vanishes. Thus the function would have to vanish except at 0 and its integral over  $\mathbb{R}$  would be 1. Such a function clearly does not exist. This has not deterred people from giving this non-existent function a name. It is called the Dirac delta function:  $\delta_0$ . One way to look at it is as the limit of the sequence of functions  $\{\delta_n\}$ , where  $\delta_n$  is the function which takes the value  $n$  on the interval  $[0, \frac{1}{n}]$  and is 0 outside this interval. Clearly  $\int_{-\infty}^{\infty} \delta_n(x) dx = 1$ . The Dirac delta function is what is known as a distribution (not to be confused with the distribution of a stochastic variable). Though the function  $\delta_0$  does not make sense, the integral  $\int_{-\infty}^{\infty} f(x)\delta_0(x)dx$  does if we define it as  $\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f(x)\delta_n(x)dx$ . To compute this integral let  $F(t) = \int_{-\infty}^t f(x)dx$ , then  $F'(x) = f(x)$  and we have  $\int_{-\infty}^{\infty} f(x)\delta_n(x)dx = n \int_0^{\frac{1}{n}} F(x)dx = n(F(\frac{1}{n}) - F(0))$ . Hence  $\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f(x)\delta_n(x)dx = \lim_{n \rightarrow \infty} \frac{F(\frac{1}{n}) - F(0)}{\frac{1}{n}} = F'(0) = f(0)$ . Thus we have  $\int_{-\infty}^{\infty} f(x)\delta_0(x)dx = f(0)$ . It follows that we have  $\phi_{W_t}(x) = \int_{-\infty}^{\infty} \phi_{W_t|W_0}(x, y)\delta_0(y)dy = \phi_{W_t|W_0}(x, 0) = \frac{1}{\sqrt{2\pi Dt}} \exp(\frac{-x^2}{2Dt})$ . Thus  $W_t \sim N(0, Dt)$ .

Now consider the increment  $W_{t_1} - W_{t_2}$ . Let  $B_z = \{(x, y) \in \mathbb{R}^2 | x - y < z\}$  then  $D_{W_{t_1} - W_{t_2}}(z) = \mu(W_{t_1} - W_{t_2} < z) = \int_{B_z} \phi_{W_{t_1}, W_{t_2}}(x, y) dx dy$ . We can rewrite this integral

$$\begin{aligned} \int_{B_z} \phi_{W_{t_1}, W_{t_2}}(x, y) dx dy &= \int_{-\infty}^{\infty} \int_{-\infty}^{y+z} \phi_{W_{t_1}|W_{t_2}}(x|y) \phi_{W_{t_2}}(y) dx dy \\ &= \frac{1}{\sqrt{2\pi D(t_1 - t_2)}} \frac{1}{\sqrt{2D\pi t_2}} \int_{-\infty}^{\infty} \int_{-\infty}^{y+z} \exp\left(\frac{-(x-y)^2}{2D(t_1 - t_2)}\right) \exp\left(\frac{-y^2}{2Dt_2}\right) dx dy \\ &= \frac{1}{\sqrt{2\pi D(t_1 - t_2)}} \frac{1}{\sqrt{2D\pi t_2}} \int_{-\infty}^{\infty} \exp\left(\frac{-y^2}{2Dt_2}\right) \int_{-\infty}^z \exp\left(\frac{-u^2}{2D(t_1 - t_2)}\right) du dy \\ &= \frac{1}{\sqrt{2\pi D(t_1 - t_2)}} \int_{-\infty}^z \exp\left(\frac{-u^2}{2D(t_2 - t_1)}\right) du \end{aligned}$$

where we have used the change of variables  $u = x - y$  and the fact that  $\frac{1}{\sqrt{2Dt_2}} \int_{-\infty}^{\infty} \exp\left(\frac{-y^2}{2Dt_2}\right) dy = 1$ . This shows that the density of the increment  $W_{t_1} - W_{t_2}$  is given by  $\phi_{W_{t_1} - W_{t_2}}(x - y) = \phi_{W_{t_1}|W_{t_2}}(x|y)$  and hence it is also normal with mean 0 and variance  $D(t_1 - t_2)$ .

Consider  $t_1 > t_2 \geq t_3 > t_4$  and the two increments  $W_{t_1} - W_{t_2}$  and  $W_{t_3} - W_{t_4}$ . We have

$$\mu(W_{t_1} - W_{t_2} < z_1, W_{t_3} - W_{t_4} < z_2) = \int_{B_{z_1}} \int_{B_{z_2}} \phi_{W_{t_1}, W_{t_2}, W_{t_3}, W_{t_4}}(x_1, x_2, x_3, x_4) dx_1 dx_2 dx_3 dx_4$$

Using the Markov property we have

$$\phi_{W_{t_1}, W_{t_2}, W_{t_3}, W_{t_4}}(x_1, x_2, x_3, x_4) = \phi_{W_{t_1}|W_{t_2}}(x_1|x_2) \phi_{W_{t_2}, W_{t_3}, W_{t_4}}(x_2, x_3, x_4)$$

Writing this out we get

$$\begin{aligned} \mu(W_{t_1} - W_{t_2} < z_1, W_{t_3} - W_{t_4} < z_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{x_4 + z_2} \int_{-\infty}^{\infty} \int_{-\infty}^{x_2 + z_1} \phi_{W_{t_1}|W_{t_2}}(x_1|x_2) \phi_{W_{t_2}, W_{t_3}, W_{t_4}}(x_2, x_3, x_4) dx_1 dx_2 dx_3 dx_4 \end{aligned}$$

Changing variables  $u = x_1 - x_2$  in the inner most integral and using that

$\phi_{W_{t_1}|W_{t_2}}(x_1|x_2) = \phi_{W_{t_1}-W_{t_2}}(u)$  we can write this as

$$\begin{aligned}
& \int_{-\infty}^{\infty} \int_{-\infty}^{x_4+z_2} \int_{-\infty}^{\infty} \int_{-\infty}^{z_1} \phi_{W_{t_1}-W_{t_2}}(u) du \phi_{W_{t_2},W_{t_3},W_{t_4}}(x_2, x_3, x_4) dx_2 dx_3 dx_4 \\
&= \mu(W_{t_1} - W_{t_2} < z_1) \int_{-\infty}^{\infty} \int_{-\infty}^{x_4+z_2} \int_{-\infty}^{\infty} \phi_{W_{t_2},W_{t_3},W_{t_4}} dx_2 dx_3 dx_4 \\
&= \mu(W_{t_1} - W_{t_2} < z_1) \int_{-\infty}^{\infty} \int_{-\infty}^{x_4+z_2} \phi_{W_{t_3},W_{t_4}} dx_3 dx_4 \\
&= \mu(W_{t_1} - W_{t_2} < z_1) \mu(W_{t_3} - W_{t_4} < z_2)
\end{aligned}$$

where we have used that  $\phi_{W_{t_3},W_{t_4}}$  is the marginal density  $\int_{-\infty}^{\infty} \phi_{W_{t_2},W_{t_3},W_{t_4}} dx_2$ . This proves that the increments are independent stochastic variables. Of course  $W_{t_1}$  and  $W_{t_2}$  themselves are not independent.

Let  $\{X_t\}_{t \in [0, T]}$  be a stochastic process in continuous time and let  $\omega \in \Omega$ . Then we have a function  $t \mapsto X_t(\omega)$ ,  $[0, T] \rightarrow \mathbb{R}$ . The graphs of these functions for different  $\omega$ 's are called the *sample paths* of the stochastic process. The sample paths are often very complicated. For instance the sample paths of the Wiener process are continuous but nowhere differentiable.

It is useful to generate sample paths of stochastic processes, both for purposes of simulation and visualization. To generate a simulated sample path of the Wiener process one can use the following procedure: divide the interval  $[0, T]$  into subintervals of length  $T/n$ . Use a random number generator to generate  $n$  samples,  $\{x_1, x_2, \dots, x_n\}$  from a normal distribution with mean 0 and variance  $D \frac{T}{n}$ . Then graph the points  $(i \frac{T}{n}, x_1 + x_2 + \dots + x_i)$ ,  $i = 0, 1, \dots, n$ . (*Exercise: Write a MATLAB program to generate sample paths for Wiener processes with different variances. Also simulate sample paths for the Cauchy process i.e. assume that the increments are independent and the density*

$\phi_{X_{t_1}-X_{t_2}}(u) = \frac{\pi}{(u^2 + (t_1 - t_2)^2)}$ . *What is the major difference between the sample paths?*)

**Definition 3.0.1** A filtration on the probability space  $(\Omega, \mathcal{F}, \mu)$  is a system of sub- $\sigma$ -algebras  $\mathcal{F}_t \subset \mathcal{F}$  indexed by  $t \in I$  such that

- $\mathcal{F}_0$  consists of all the measurable sets of measure 0 or 1
- $t_2 \leq t_1$  implies  $\mathcal{F}_{t_2} \subset \mathcal{F}_{t_1}$
- In the case when  $I$  is a continuous interval we require that  $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ .

We stress that this condition does not apply when  $I$  is a finite set

**Definition 3.0.2** Let  $\{X_t\}_{t \in I}$  be a stochastic process and  $\{\mathcal{F}_t\}$  a filtration. The process is said to be adapted to the filtration if for all  $t \in I$ ,  $X_t$  is measurable with respect to  $\mathcal{F}_t$

Remark that a stochastic process  $\{X_t\}$  determines a filtration by letting  $\mathcal{G}_t$  be the  $\sigma$ -algebra generated by the stochastic processes  $\{X_s\}_{s \leq t}$ . So stochastic processes and filtrations are naturally related. Given a filtration  $\{\mathcal{F}_t\}_{t \in I}$ ,  $\{X_t\}$  is adapted if and only if  $\mathcal{G}_t \subset \mathcal{F}_t$  for all  $t$ .

**Lemma 3.0.4** Let  $X$  be a random variable. Then  $\mathbb{E}(X|\mathcal{F}_0) = \mathbb{E}(X)$

Since a constant function is measurable with respect to any  $\sigma$ -algebra,  $\mathbb{E}(X)$  is measurable with respect to  $\mathcal{F}_0$ .

Let  $A \in \mathcal{F}_0$ , so  $\mu(A) = 0$  or 1. If  $\mu(A) = 0$  we have  $\int_A X d\mu = 0 = \int_A \mathbb{E}(X) d\mu$ . If  $\mu(A) = 1$  we have  $\mu(\Omega - A) = 0$  so  $\int_A X d\mu = \int_\Omega X d\mu = \mathbb{E}(X) = \int_A \mathbb{E}(X) d\mu$ . This proves that  $\mathbb{E}(X|\mathcal{F}_0) = \mathbb{E}(X)$  a.e.

**Definition 3.0.3** Let  $\{\mathcal{F}_t\}$  be a filtration and  $\{X_t\}$  an adapted process. For  $t_1 \geq t_2$  we can consider  $\mathbb{E}(X_{t_1}|\mathcal{F}_{t_2})$ . This is a random variable with respect to  $\mathcal{F}_{t_2}$ . The process  $\{X_t\}$  is said to be a martingale with respect to the filtration  $\{\mathcal{F}_t\}$  if for all  $t_1, t_2$  with  $t_1 \geq t_2$  we have  $\mathbb{E}(X_{t_1}|\mathcal{F}_{t_2}) = X_{t_2}$

### 3.1 Discrete Time Processes

We shall consider in more detail the discrete time case where the index set is finite, thus we have a sequence of stochastic variables  $X_0, X_1, \dots, X_T$ . We can often gain information about the continuous time case by approximating with discrete time processes, we have already seen an example of this in the simulation of sample paths. Here we shall develop some aspects of stochastic calculus in the discrete time case.

Most of the following is taken from Dothan's book: Prices in Financial Markets. I highly recommend this book though it will not be used directly in the course.

**Definition 3.1.1** Let  $(\Omega, \mathcal{F}, \mu, \{\mathcal{F}_t\}_{t=0,1,\dots,T})$  be a filtered probability space. A stochastic process  $\{X_t\}_{t=0,1,\dots,T}$  is said to be predictable if  $X_t$  is measurable with respect to  $\mathcal{F}_{t-1}$ .

**Proposition 3.1.1** A predictable martingale is constant

Let  $\{X_t\}$  be a predictable martingale. Then we have  $\mathbb{E}(X_t|\mathcal{F}_{t-1}) = X_{t-1}$ . On the other hand since  $X_t$  is measurable with respect to  $\mathcal{F}_{t-1}$  we have  $\mathbb{E}(X_t|\mathcal{F}_{t-1}) = X_t$  so  $X_t = X_{t-1}$  for all  $t$ .

**Theorem 3.1.1** (The Doob Decomposition) Let  $\{X_t\}$  be an adapted process. Given a real number  $a$  there exists a unique decomposition

$$X_t = A_t + W_t$$

where  $\{A_t\}$  is a predictable process such that  $A_0 = a$  and  $\{W_t\}$  is a martingale

Define

$$A_t = \begin{cases} a & t = 0 \\ A_{t-1} - X_{t-1} + \mathbb{E}(X_t|\mathcal{F}_{t-1}) & t \geq 1 \end{cases}$$

and

$$W_t = \begin{cases} X_0 - a & t = 0 \\ W_{t-1} + X_t - \mathbb{E}(X_t|\mathcal{F}_{t-1}) & t \geq 1 \end{cases}$$

We have  $A_t + W_t = A_{t-1} - X_{t-1} + \mathbb{E}(X_t|\mathcal{F}_{t-1}) + W_{t-1} + X_t - \mathbb{E}(X_t|\mathcal{F}_{t-1}) = A_{t-1} + W_{t-1} - X_{t-1} + X_t = X_t$

It is clear that  $A_t$  is measurable with respect to  $\mathcal{F}_{t-1}$  so  $\{A_t\}$  is a predictable process. Also we have  $\mathbb{E}(W_t|\mathcal{F}_{t-1}) = W_{t-1} - \mathbb{E}(X_t|\mathcal{F}_{t-1}) + \mathbb{E}(X_t|\mathcal{F}_{t-1}) = W_{t-1}$  which shows that  $\{W_t\}$  is a martingale.

To prove uniqueness assume that  $X_t = A'_t + W'_t$  where  $\{A'_t\}$  is predictable and  $\{W'_t\}$  is a martingale. Then we have  $A_t - A'_t = W_t - W'_t$  so  $A_t - A'_t$  is both predictable and a martingale and hence by proposition 3.1.1 is constant. Since  $A_0 - A'_0 = a - a = 0$  we have  $A_t = A'_t$  for all  $t$  and so also  $W_t = W'_t$

**Example 3.2** Consider the process of the price of a stock  $\{S_t\}$  and let  $R_t = \frac{S_t - S_{t-1}}{S_{t-1}} = \frac{S_t}{S_{t-1}} - 1$ , i.e. the process of returns. We have  $W_t = W_{t-1} + R_t - \mathbb{E}(R_t|\mathcal{F}_{t-1})$  and  $\{W_t\}$  is a martingale. We get  $R_t = \mathbb{E}(R_t|\mathcal{F}_{t-1}) + W_t - W_{t-1}$ . Put  $\mu_t = \mathbb{E}(R_t|\mathcal{F}_{t-1})$  so  $R_t = \mu_t + W_t - W_{t-1}$ .

In our binomial tree example we have  $\mu_t = \mathbb{E}(\frac{S_t}{S_{t-1}}|\mathcal{F}_{t-1}) - 1 = \frac{u+d}{2} - 1$  and  $W_t - W_{t-1}$  is the stochastic variable

$$(W_t - W_{t-1})(\omega) = \frac{S_t}{S_{t-1}}(\omega) - \frac{u+d}{2} = \begin{cases} \frac{u-d}{2} & \omega_t = 1 \\ \frac{-(u-d)}{2} & \omega_t = 0 \end{cases}$$

Thus  $\{W_t - W_{t-1}\}$  is a random walk, where at time  $t$  we either move up by  $\frac{u-d}{2}$  or down by the same amount, each move having probability  $\frac{1}{2}$

From now on we shall denote stochastic processes by the simpler notation  $X, Y$  etc.

To motivate the next definition consider trading in a single stock at discrete time intervals  $t = 1, 2, \dots, T$ . Let the stock price be given by the stochastic process  $S$ . Let  $\theta_s$  denote the number of shares held in the time period  $s, s+1$ . The stochastic process  $\theta$  is called a trading strategy. The trading gain in the period  $s-1, s$  is  $\theta_{s-1}(S_s - S_{s-1})$ . Thus the total gain from the trading strategy  $\theta$  at time  $t$  is given by  $\sum_{s=1}^t \theta_{s-1}(S_s - S_{s-1})$

**Definition 3.1.2** Let  $X$  and  $Y$  be stochastic processes. We define the stochastic integral

$$\int_0^t Y dX = \begin{cases} 0 & t = 0, \\ \sum_{s=1}^t Y_{s-1}(X_s - X_{s-1}) & t > 0 \end{cases}$$

The stochastic integral  $\int_0^t Y dX$  is sometimes denoted  $Y \bullet X$  and is called the transform of the process  $X$  by the process  $Y$ .

For any process  $X$  we define the backwards difference  $\Delta X_t = X_t - X_{t-1}$  for  $t > 0$  and 0 when  $t = 0$ . Then we have  $\Delta(Y \bullet X)_t = Y_{t-1} \Delta X_t$ . This follows trivially from the definition: indeed it is clear for  $t = 0$  and for  $t > 0$  we have  $\Delta(Y \bullet X)_t = \sum_{s=1}^t Y_{s-1}(X_s - X_{s-1}) - \sum_{s=1}^{t-1} Y_{s-1}(X_s - X_{s-1}) = Y_{t-1}(X_t - X_{t-1})$

The next theorem plays a fundamental role in the theory

**Theorem 3.1.2** *Let  $Y$  be an adapted stochastic process and  $X$  a martingale. Then the stochastic integral  $\int YdX$  is a martingale*

The proof of this in the discrete case is very simple: we have  $\mathbb{E}(\int_0^t YdX|\mathcal{F}_{t-1}) = \sum_{s=1}^t \mathbb{E}(Y_{s-1}(X_s - X_{s-1})|\mathcal{F}_{t-1})$ . Since  $Y_{s-1}(X_s - X_{s-1})$  is measurable with respect to  $\mathcal{F}_{t-1}$  for  $s < t$  we have  $\mathbb{E}(Y_{s-1}(X_s - X_{s-1})|\mathcal{F}_{t-1}) = Y_{s-1}(X_s - X_{s-1})$  for  $s < t$ .

Now  $\mathbb{E}(Y_{t-1}(X_t - X_{t-1})|\mathcal{F}_{t-1}) = \mathbb{E}(Y_{t-1}X_t|\mathcal{F}_{t-1}) - \mathbb{E}(Y_{t-1}X_{t-1}|\mathcal{F}_{t-1})$ . Since  $Y_{t-1}X_{t-1}$  is measurable with respect to  $\mathcal{F}_{t-1}$  it follows that  $\mathbb{E}(Y_{t-1}X_{t-1}|\mathcal{F}_{t-1}) = Y_{t-1}X_{t-1}$ . Again because  $Y_{t-1}$  is measurable with respect to  $\mathcal{F}_{t-1}$  we have by theorem 2.4.2 that  $\mathbb{E}(Y_{t-1}X_t|\mathcal{F}_{t-1}) = Y_{t-1}\mathbb{E}(X_t|\mathcal{F}_{t-1})$  and since  $X$  is a martingale  $\mathbb{E}(X_t|\mathcal{F}_{t-1}) = X_{t-1}$ .

Thus  $\mathbb{E}(Y_{t-1}X_t|\mathcal{F}_{t-1}) = Y_{t-1}X_{t-1}$  and hence  $\mathbb{E}(Y_{t-1}(X_t - X_{t-1})|\mathcal{F}_{t-1}) = 0$ . Putting this together we get

$$\begin{aligned} \mathbb{E}(\int_0^t YdX|\mathcal{F}_{t-1}) &= \mathbb{E}(Y_{t-1}(X_t - X_{t-1})|\mathcal{F}_{t-1}) + \sum_{s=1}^{t-1} \mathbb{E}(Y_{s-1}(X_s - X_{s-1})|\mathcal{F}_{t-1}) = \\ &= \sum_{s=1}^{t-1} Y_{s-1}(X_s - X_{s-1}) = \int_0^{t-1} YdX, \text{ which proves the martingale property.} \end{aligned}$$

To put the stochastic integral in perspective we can define for a pair of functions  $f$  and  $g$ , the Riemann-Stieltjes integral  $\int_a^b f dg$  as follows: we divide the interval  $[a, b]$  into subintervals  $[t_i, t_{i+1}]$  of length  $\Delta t$  and consider the Riemann sums  $\sum f(t_i)(g(t_{i+1}) - g(t_i))$ .  $\int_a^b f dg$  is defined by  $\lim_{\Delta t \rightarrow 0} \sum f(t_i)(g(t_{i+1}) - g(t_i))$  if this limit exists. If  $g$  satisfies some mild conditions (like bounded variation), the limit is independent of the point where in the interval  $[t_i, t_{i+1}]$ ,  $f$  is evaluated (here at the left endpoint of the subinterval,  $t_i$ ).

The stochastic integral in the continuous case is constructed in similar fashion as a limit of Riemann sums (in the discrete case the stochastic integral is the Riemann sum). However the function  $t \mapsto X_t(\omega)$ , except for trivial cases is not of bounded variation and so the limit does depend on the choice of the point at which  $Y$  is evaluated. Thus by taking other points in the interval  $[t_i, t_{i+1}]$  such as  $\alpha t_i + (1 - \alpha)t_{i+1}$  and evaluating at these points, one can construct other stochastic integrals (they are called *Stratonovic integrals*, while the stochastic integral we are considering is called the *Ito integral*). The Stratonovic integral, however does not preserve martingales and furthermore, by looking at values at a point other than  $t_i$  we are in some sense looking into the future to get information about the present. This is one reason why



it is the Ito integral that is used almost exclusively in financial applications.

**Definition 3.1.3** Let  $X$  and  $Y$  be stochastic processes. We define the optional quadratic covariation process,  $[X, Y]$  by

$$[X, Y]_t = \begin{cases} X_0 Y_0 & t = 0 \\ X_0 Y_0 + \sum_{s=1}^t \Delta X_s \Delta Y_s & t \geq 1 \end{cases}$$

The process  $[X, X]$  is called the optional quadratic variation process of  $X$ .

Remark that the optional quadratic covariation process does not depend on the filtration  $\{\mathcal{F}_t\}$

We define the predictable quadratic covariation process,  $\langle X, Y \rangle$  by

$$\langle X, Y \rangle_t = \begin{cases} \mathbb{E}(X_0 Y_0) & t = 0 \\ \mathbb{E}(X_0 Y_0) + \sum_{s=1}^t \mathbb{E}(\Delta X_s \Delta Y_s | \mathcal{F}_{s-1}) & t > 0 \end{cases}$$

The process  $\langle X, X \rangle$  is called the predictable quadratic variation process of  $X$ .

Remark that  $\mathbb{E}(\Delta X_s \Delta Y_s | \mathcal{F}_{s-1})$  is measurable with respect to  $\mathcal{F}_{t-1}$  for  $s = 1, \dots, t$ , hence the predictable quadratic covariation process is in fact a predictable process

**Example 3.3** Consider the random variable  $w_t$  on the probability space  $\Omega_T$

defined by  $w_t(\omega) = \begin{cases} D & \omega_t = 1 \\ -D & \omega_t = 0 \end{cases}$  and let  $W$  be the stochastic process defined

$$\text{by } W_t = \begin{cases} 0 & t = 0 \\ \sum_{s=1}^t w_s & t \geq 1 \end{cases}$$

The random variable  $W_t$  takes the values  $\{(t-2i)D\}_{i=0,1,\dots,t}$  with  $\mu(W_t = (t-2i)D) = \binom{t}{i} 2^{-t}$ . This stochastic process is called a symmetric random walk. We have

$$\mathbb{E}(W_t) = \sum_{i=0}^t \binom{t}{i} 2^{-t} (t-2i)D = \sum_{i=0}^t \binom{t}{i} 2^{-t} ((t-2i) + (t-2(t-i)))D = 0$$

Since the  $w_t$ 's are independent, we can compute the variance of  $W_t$  as the sum of the variances of the  $w_s$ 's,  $\text{Var}(W_t) = \sum_{s=1}^t \text{Var}(w_s)$  and  $\text{Var}(w_s) = D^2$  and so  $\text{Var}(W_t) = tD^2$ .

The optional quadratic variation process  $[W, W]$  is given by  $[W, W]_t = \sum_{s=0}^t (\Delta W_s)^2 = \sum_{s=0}^t w_s^2 = tD^2$  and the predictable quadratic variation process  $\langle W, W \rangle$  is given by  $\langle W, W \rangle_t = \sum_{s=0}^t \mathbb{E}((\Delta W_s)^2 | \mathcal{F}_{s-1}) = tD^2$  because  $w_s^2$  is constant  $= D^2$  and hence is measurable with respect to  $\mathcal{F}_{s-1}$  so  $\mathbb{E}(w_s^2 | \mathcal{F}_{s-1}) = w_s^2 = D^2$

Let  $\eta \in \Omega_{t-1}$  then we have  $\int_{U_\eta} W_t d\mu = \int_{U_\eta} W_{t-1} + \int_{U_\eta} w_t d\mu$ . Now  $\int_{U_\eta} w_t d\mu = D\mu(U_{\eta,1}) - D\mu(U_{\eta,0}) = 0$ . It follows that  $\int_{U_\eta} W_t d\mu = \int_{U_\eta} W_{t-1} d\mu$  so  $\mathbb{E}(W_t | \mathcal{F}_{t-1}) = W_{t-1}$ , in other words  $W$  is a martingale.

**Theorem 3.1.3** *If  $X$  and  $Y$  are martingales then  $XY - [X, Y]$  and  $XY - \langle X, Y \rangle$  are martingales*

Clearly  $[X, Y]_t = \Delta X_t \Delta Y_t + [X, Y]_{t-1}$  and so  $\mathbb{E}(X_t Y_t - [X, Y]_t | \mathcal{F}_{t-1}) = \mathbb{E}(X_t Y_t - \Delta X_t \Delta Y_t | \mathcal{F}_{t-1}) - [X, Y]_{t-1}$  since  $[X, Y]_{t-1}$  is measurable with respect to  $\mathcal{F}_{t-1}$ . Now  $X_t Y_t - \Delta X_t \Delta Y_t = X_t Y_t - (X_t - X_{t-1})(Y_t - Y_{t-1}) = X_t Y_{t-1} + X_{t-1} Y_t - X_{t-1} Y_{t-1}$  so  $\mathbb{E}(X_t Y_t - \Delta X_t \Delta Y_t | \mathcal{F}_{t-1}) = \mathbb{E}(X_t Y_{t-1} | \mathcal{F}_{t-1}) + \mathbb{E}(X_{t-1} Y_t | \mathcal{F}_{t-1}) - X_{t-1} Y_{t-1}$ . But since  $X_{t-1}$  and  $Y_{t-1}$  are measurable with respect to  $\mathcal{F}_{t-1}$  we get by theorem 2.4.2 that  $\mathbb{E}(X_t Y_{t-1} | \mathcal{F}_{t-1}) = Y_{t-1} \mathbb{E}(X_t | \mathcal{F}_{t-1}) = Y_{t-1} X_{t-1}$ , the last equality because  $X$  is a martingale. Similarly  $\mathbb{E}(X_{t-1} Y_t | \mathcal{F}_{t-1}) = X_{t-1} Y_{t-1}$ . It follows that  $\mathbb{E}(X_t Y_t - \Delta X_t \Delta Y_t | \mathcal{F}_{t-1}) = X_{t-1} Y_{t-1}$  and so  $\mathbb{E}(X_t Y_t - [X, Y]_t | \mathcal{F}_{t-1}) = X_{t-1} Y_{t-1} - [X, Y]_{t-1}$ . This proves the first case.

For the second case we have  $\mathbb{E}(X_t Y_t - \langle X, Y \rangle_t) = \mathbb{E}(X_t Y_t - \mathbb{E}(\Delta X_t \Delta Y_t | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}) - \langle X, Y \rangle_{t-1}$ . As above we get  $\mathbb{E}(\Delta X_t \Delta Y_t | \mathcal{F}_{t-1}) = \mathbb{E}(X_t Y_t | \mathcal{F}_{t-1}) - X_{t-1} Y_{t-1}$ . Hence  $\mathbb{E}(X_t Y_t - \Delta X_t \Delta Y_t | \mathcal{F}_{t-1}) = \mathbb{E}(X_t Y_t | \mathcal{F}_{t-1}) - \mathbb{E}(\mathbb{E}(X_t Y_t | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}) + X_{t-1} Y_{t-1} = X_{t-1} Y_{t-1}$  because by definition of conditional expectation  $\mathbb{E}(X_t Y_t | \mathcal{F}_{t-1})$  is measurable with respect to  $\mathcal{F}_{t-1}$  so  $\mathbb{E}(\mathbb{E}(X_t Y_t | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}) = \mathbb{E}(X_t Y_t | \mathcal{F}_{t-1})$  and hence the two first terms cancel. This proves the second statement.

Applying this result to the symmetric random walk  $W$ , we get that the process  $\{W_t^2 - tD^2\}_t$  is a martingale. Remark that  $W^2$  itself is not a martingale. Indeed  $\int_{U_\eta} W_t^2 d\mu = \int_{U_\eta} (W_{t-1}^2 + w_t^2 + 2W_{t-1} w_t) d\mu = \int_{U_\eta} W_{t-1}^2 d\mu + D^2 \mu(U_\eta) + W_{t-1}(\eta) \int_{U_\eta} w_t d\mu = \int_{U_\eta} W_{t-1}^2 d\mu + D^2 \mu(U_\eta)$  because  $W_{t-1}$  is constant on  $U_\eta$  and  $\int_{U_\eta} w_t d\mu = 0$

**Theorem 3.1.4** *Let  $A, X, Y$  be stochastic processes, then*

$$[A \bullet X, Y] = A \bullet [X, Y]$$

and

$$\langle A \bullet X, Y \rangle = A \bullet \langle X, Y \rangle$$

The proof of this is straight forward. For the first equality to be proved, the left hand side is

$$[A \bullet X, Y]_t = (A \bullet X)_0 Y_0 + \sum_{s=1}^t \Delta(A \bullet X)_s \Delta Y_s = \sum_{s=1}^t A_{s-1} \Delta X_s \Delta Y_s$$

The right hand side is

$$(A \bullet [X, Y])_t = \sum_{s=1}^t A_s \Delta[X, Y]_s = \sum_{s=1}^t A_{s-1} \Delta X_s \Delta Y_s$$

The last equality follows since

$$\begin{aligned} \Delta[X, Y]_s &= [X, Y]_s - [X, Y]_{s-1} \\ &= X_0 Y_0 + \sum_{u=1}^s (X_u - X_{u-1})(Y_u - Y_{u-1}) - (X_0 Y_0 + \sum_{u=1}^{s-1} (X_u - X_{u-1})(Y_u - Y_{u-1})) \\ &= (X_s - X_{s-1})(Y_s - Y_{s-1}) \\ &= \Delta X_s \Delta Y_s \end{aligned}$$

For the second equality, the left hand side is

$$\begin{aligned} \langle A \bullet X, Y \rangle_t &= \mathbb{E}((A \bullet X)_0 Y_0) + \sum_{s=1}^t \mathbb{E}(\Delta(A \bullet X)_s \Delta Y_s | \mathcal{F}_{s-1}) \\ &= \sum_{s=1}^t \mathbb{E}(A_{s-1} \Delta X_s \Delta Y_s | \mathcal{F}_{s-1}) \\ &= \sum_{s=1}^t A_{s-1} \mathbb{E}(\Delta X_s \Delta Y_s | \mathcal{F}_{s-1}) \end{aligned}$$

The last equation follows from theorem 2.4.2

The right hand side is

$$\begin{aligned} (A \bullet \langle X, Y \rangle)_t &= \sum_{s=1}^t A_{s-1} \Delta \langle X, Y \rangle_s \\ &= \sum_{s=1}^t A_{s-1} \mathbb{E}(\Delta X_s \Delta Y_s | \mathcal{F}_{s-1}) \end{aligned}$$

the last equality follows from the definition, since  $\Delta\langle X, Y \rangle_s = \langle X, Y \rangle_s - \langle X, Y \rangle_{s-1} = \mathbb{E}((X_s - X_{s-1})(Y_s - Y_{s-1}) | \mathcal{F}_{t-1}) = \mathbb{E}(\Delta X_s \Delta Y_s | \mathcal{F}_{t-1})$

**Proposition 3.1.2**  $\int_0^t X dX = \frac{1}{2}X_t^2 - \frac{1}{2}[X, X]_t$

Remark that except for the adjustment term  $\frac{1}{2}[X, X]_t$ , this is the usual formula for  $\int_0^t X dX$ .

The proof is a straight forward computation using the identity  $b(a - b) = \frac{1}{2}a^2 - \frac{1}{2}b^2 - \frac{1}{2}(a - b)^2$ . Indeed

$$\begin{aligned} \int_0^t X dX &= \sum_{s=1}^t X_{s-1}(X_s - X_{s-1}) \\ &= \sum_{s=1}^t \frac{1}{2}(X_s^2 - X_{s-1}^2) - \sum_{s=1}^t \frac{1}{2}(X_s - X_{s-1})^2 \\ &= \frac{1}{2}X_t^2 - \frac{1}{2}[X, X]_t \end{aligned}$$

**Example 3.4**  $\int_0^t W dW = \frac{1}{2}W_t^2 - \frac{1}{2}[W, W]_t = \frac{1}{2}(W_t^2 - tD^2)$ . This gives another proof that  $\{W_t^2 - tD^2\}_t$  is a martingale

Remark that if we use the Stratonovic integral and take the evaluation point to be  $\frac{X_s + X_{s-1}}{2}$  then the usual formula for  $\int_0^t X dX = \frac{1}{2}X_t^2 - \frac{1}{2}X_0^2$  holds, but this integral no longer preserves the martingale property (*Exercise: Verify this*).

We also have the integration by parts formula

**Proposition 3.1.3**

$$\int_0^t X dY = X_t Y_t - \int_0^t Y dX - [X, Y]_t$$

Indeed  $[X, Y] = \frac{1}{2}([X + Y, X + Y] - [X, X] - [Y, Y])$  and hence by the previous proposition we get

$$\begin{aligned} [X, Y]_t &= \frac{1}{2}(X + Y)_t^2 - \int_0^t (X + Y)d(X + Y) - \frac{1}{2}X_t^2 + \int_0^t X dX - \frac{1}{2}Y_t^2 + \int_0^t Y dY \\ &= X_t Y_t - \int_0^t X dY - \int_0^t Y dX \end{aligned}$$

Instead of writing the integral one often uses a short hand notation as a differential, instead of writing  $A_t = \int_0^t Y dX$  one would write  $dA_t = Y_t dX_t$ . Such an expression is called a stochastic differential equation (SDE). In this notation the formula above would become similar to the familiar formula for the derivative of a product:  $d(X_t Y_t) = X_t dY_t + Y_t dX_t + d[X, Y]_t$ . Also we can rewrite the definition of the optional quadratic covariation process  $d[X, Y]_t = dX_t dY_t$  and the predictable covariation process  $d\langle X, Y \rangle_t = \mathbb{E}(dX_t dY_t | \mathcal{F}_{t-1})$ . We also get the following formula for the quadratic variation process of the symmetric random walk with jumps  $\pm D$ ,  $W$ :  $(dW_t)^2 = d[W, W]_t = D^2 dt$  or  $dW_t = D\sqrt{dt}$

**Example 3.5** Consider the stochastic differential equation,  $dX = X dY$ . If  $Y$  is an adapted process and  $x_0$  is an arbitrary constant then there exists a unique solution  $X$  with  $X_0 = x_0$ . The process  $X$  is given by

$$X_t = x_0 \prod_{s=1}^t (1 + \Delta Y_s)$$

Indeed the stochastic differential equation is shorthand notation for integral equation

$$X_t = x_0 + \int_0^t X dY$$

which we can solve recursively

$$X_1 = x_0 + \int_0^1 X dY = x_0 + X_0 \Delta Y_1 = x_0(1 + \Delta Y_1)$$

$$\begin{aligned} X_2 &= x_0 + \int_0^2 X dY \\ &= x_0 + X_0 \Delta Y_1 + X_1 \Delta Y_2 \\ &= x_0 + x_0 \Delta Y_1 + x_0(1 + \Delta Y_1) \Delta Y_2 \\ &= x_0(1 + \Delta Y_1 + (1 + \Delta Y_1) \Delta Y_2) \\ &= x_0(1 + \Delta Y_1)(1 + \Delta Y_2) \end{aligned}$$

and so on.

The process  $X$  with  $X_0 = 1$  solving the SDE  $dX = X dY$  is called the Stochastic Exponential of the process  $Y$  and is denoted  $\mathcal{E}(Y)$ . Remark that if  $Y$  is a martingale then  $\mathcal{E}(Y)$  is also a martingale. (Exercise: Show that for two processes  $A, B$  we have  $\mathcal{E}(A)\mathcal{E}(B) = \mathcal{E}(A + B + [A, B])$ )

Finally we shall discuss what happens to the martingale property under a change of measure.

Assume that  $\nu$  is another probability measure on  $(\Omega, \mathcal{F})$  and assume  $\nu$  is absolutely continuous with respect to  $\mu$  i.e.  $\mu(A) = 0 \Rightarrow \nu(A) = 0$ . By the Radon-Nikodym theorem there exists an integrable function  $\rho$  such that for all  $B \in \mathcal{F}$ ,  $\nu(B) = \int_B \rho d\mu$ . The Radon-Nikodym derivative  $\rho = \frac{d\nu}{d\mu}$  is a random variable with respect to  $\mathcal{F}$  and hence we can consider the stochastic process  $Z_t = \mathbb{E}\left(\frac{d\nu}{d\mu} \middle| \mathcal{F}_t\right)$ . This process is called the *Likelihood Ratio Process*. Remark that  $Z$  is clearly a martingale.

To distinguish between expectations with respect to the two measures we shall use the notation  $\mathbb{E}_\mu$  and  $\mathbb{E}_\nu$ .

**Theorem 3.1.5** *Let  $X$  be an adapted process then for any  $1 \leq s \leq t \leq T$  we have*

$$\mathbb{E}_\mu(Z_t X_t | \mathcal{F}_s) = \mathbb{E}_\mu(Z_t | \mathcal{F}_s) \mathbb{E}_\nu(X_t | \mathcal{F}_s)$$

Since  $Z$  is a martingale we have to show  $\mathbb{E}_\mu(Z_t X_t | \mathcal{F}_s) = Z_s \mathbb{E}_\nu(X_t | \mathcal{F}_s)$ . Now  $\mathbb{E}_\mu(Z_t X_t | \mathcal{F}_s)$  is characterized by  $\int_B Z_t X_t d\mu = \int_B \mathbb{E}_\mu(Z_t X_t | \mathcal{F}_s) d\mu$  for any  $B \in \mathcal{F}_s$  and  $\mathbb{E}_\mu(Z_t X_t | \mathcal{F}_s)$  is measurable with respect to  $\mathcal{F}_s$ .

Remark that for all  $B \in \mathcal{F}_s$  we have  $\nu(B) = \int_B Z_s d\mu$  hence  $Z_s$  is the Radon-Nikodym derivative  $\frac{d\nu}{d\mu}$  where the measures are restricted to the  $\sigma$ -algebra  $\mathcal{F}_s$

By theorem 1.5.4 we have  $\int_B Z_t X_t d\mu = \int_B X_t d\nu$  and  $\int_B X_t d\nu = \int_B \mathbb{E}_\nu(X_t | \mathcal{F}_s) d\nu$ . Again by theorem 1.5.4 we have  $\int_B \mathbb{E}_\nu(X_t | \mathcal{F}_s) d\nu = \int_B Z_s \mathbb{E}_\nu(X_t | \mathcal{F}_s) d\mu$ . Putting these identities together we get  $\int Z_t X_t d\mu = \int_B Z_s \mathbb{E}_\nu(X_t | \mathcal{F}_s)$  for all  $B \in \mathcal{F}_s$ . Since  $Z_s \mathbb{E}_\nu(X_t | \mathcal{F}_s)$  is measurable with respect to  $\mathcal{F}_s$  this shows that  $\mathbb{E}_\mu(Z_t X_t | \mathcal{F}_s) = Z_s \mathbb{E}_\nu(X_t | \mathcal{F}_s)$

**Corollary 3.1.1** *An adapted process  $X$  is a martingale with respect to  $\nu$  if and only if  $ZX$  is a martingale with respect to  $\mu$*

*Exercise: Prove this*

**Theorem 3.1.6** (*Girsanov's Theorem*) *Assume  $X$  is a martingale with respect to  $\mu$  then the stochastic process  $\{X_t - \int_0^t \frac{d(X,Z)}{Z}\}_t$  is a martingale with respect to  $\nu$*

The conditional expectation

$$\mathbb{E}_\nu(X_t - \int_0^t \frac{d\langle X, Z \rangle}{Z} | \mathcal{F}_{t-1})$$

is equal to

$$\mathbb{E}_\nu(X_t | \mathcal{F}_{t-1}) - \int_0^{t-1} \frac{d\langle X, Z \rangle}{Z} - \mathbb{E}_\nu\left(\frac{\langle X, Z \rangle_t - \langle X, Z \rangle_{t-1}}{Z_{t-1}} | \mathcal{F}_{t-1}\right)$$

Since  $\langle X, Z \rangle$  is a predictable process  $\frac{\langle X, Z \rangle_t - \langle X, Z \rangle_{t-1}}{Z_{t-1}}$  is measurable with respect to  $\mathcal{F}_{t-1}$  and so

$$\mathbb{E}_\nu\left(\frac{\langle X, Z \rangle_t - \langle X, Z \rangle_{t-1}}{Z_{t-1}} | \mathcal{F}_{t-1}\right) = \frac{\langle X, Z \rangle_t - \langle X, Z \rangle_{t-1}}{Z_{t-1}}$$

Now

$$\begin{aligned} \frac{\langle X, Z \rangle_t - \langle X, Z \rangle_{t-1}}{Z_{t-1}} &= \frac{\mathbb{E}_\mu((X_t - X_{t-1})(Z_t - Z_{t-1}) | \mathcal{F}_{t-1})}{Z_{t-1}} \\ &= \frac{\mathbb{E}_\mu((X_t - X_{t-1})(Z_t - Z_{t-1}) | \mathcal{F}_{t-1})}{\mathbb{E}_\mu(Z_t | \mathcal{F}_{t-1})} \\ &= \frac{\mathbb{E}_\mu(Z_t(X_t - X_{t-1})(1 - \frac{Z_{t-1}}{Z_t}) | \mathcal{F}_{t-1})}{\mathbb{E}_\mu(Z_t | \mathcal{F}_{t-1})} \\ &= \mathbb{E}_\nu((X_t - X_{t-1})(1 - \frac{Z_{t-1}}{Z_t}) | \mathcal{F}_{t-1}) \end{aligned}$$

where the second equality holds because  $Z$  is a martingale with respect to  $\mu$  and the last equality holds by theorem 3.1.5. We have

$$\begin{aligned} \mathbb{E}_\nu((X_t - X_{t-1})(1 - \frac{Z_{t-1}}{Z_t}) | \mathcal{F}_{t-1}) &= \mathbb{E}_\nu(X_t | \mathcal{F}_{t-1}) - X_{t-1} + Z_{t-1} \mathbb{E}_\nu(\frac{X_t - X_{t-1}}{Z_t} | \mathcal{F}_{t-1}) \\ &= \mathbb{E}_\nu(X_t | \mathcal{F}_{t-1}) - X_{t-1} + \mathbb{E}_\mu(X_t - X_{t-1} | \mathcal{F}_{t-1}) \\ &= \mathbb{E}_\nu(X_t | \mathcal{F}_{t-1}) - X_{t-1} \end{aligned}$$

The second equality follows from theorem 3.1.5 and the last equality holds because  $X$  is a  $\mu$ -martingale so  $\mathbb{E}_\mu(X_t - X_{t-1} | \mathcal{F}_{t-1}) = 0$ .

Putting this together we get

$$\begin{aligned}\mathbb{E}_\nu(X_t - \int_0^t \frac{d\langle X, Z \rangle}{Z} | \mathcal{F}_{t-1}) &= \mathbb{E}_\nu(X_t | \mathcal{F}_{t-1}) - \int_0^{t-1} \frac{d\langle X, Z \rangle}{Z} - (\mathbb{E}_\nu(X_t | \mathcal{F}_{t-1}) - X_{t-1}) \\ &= X_{t-1} - \int_0^{t-1} \frac{d\langle X, Z \rangle}{Z}\end{aligned}$$

which proves the martingale property

**Example 3.6** Consider a market with two securities:

- a risk free bond with a constant rate of interest  $r$  so the price  $B_t$  at time  $t$  is given by  $(1+r)^t$
- a stock whose price at time  $t$  is  $S_t$

$S$  is a stochastic process on some state space  $(\Omega, \mathcal{F}, \mu)$  equipped with an information structure  $\mathcal{F}$ .

Recall that we put  $\mu_t = \mathbb{E}(R_t | \mathcal{F}_{t-1})$ , the conditional expectation of the return  $R_t = \frac{S_t - S_{t-1}}{S_{t-1}}$ . Using the Doob decomposition we can write  $R_t = \mu_t + W_t - W_{t-1}$  where  $W$  is a martingale, hence  $S_t = S_{t-1}(1 + \mu_t + W_t - W_{t-1})$  or in shorthand notation  $\Delta S_t = S_{t-1}\mu_t \Delta t + S_{t-1}\Delta W_t$ , where  $\Delta t = 1$ . By recursion we get  $S_t = S_0 \prod_{s=1}^t (1 + \mu_s + W_s - W_{s-1})$ .

A measure  $\nu$  on  $(\Omega, \mathcal{F})$  is said to be an Equilibrium Price Measure if it is absolutely continuous with respect to  $\mu$  and the discounted price process  $\frac{S_t}{(1+r)^t}$  is a martingale with respect to  $\nu$ . The existence of equilibrium price measures are important for calculating prices of derivative securities.

Consider a trading strategy  $(\theta, \phi)$  i.e.  $\theta$  and  $\phi$  are predictable processes. We can then consider the process  $C_t = \theta_t S_t + \phi_t B_t$ . Thus  $C_t$  is the value of a portfolio at time  $t$  consisting of  $\theta_t$  shares of the stock and  $\phi_t$  bonds. At time  $t$  we need to buy  $\theta_{t+1} - \theta_t$  shares and  $\phi_{t+1} - \phi_t$  bonds. Thus the cost of the trading strategy at time  $t$  is  $(\theta_{t+1} - \theta_t)S_t + (\phi_{t+1} - \phi_t)B_t$ . The trading strategy is said to be self-financing if for all  $t$ , the cost at time  $t$  is 0.

**Theorem 3.1.7** If  $(\theta, \phi)$  is a self-financing trading strategy, then the discounted value process  $\frac{C_t}{(1+r)^t}$  is a martingale with respect to any price equilibrium measure



Indeed let  $\nu$  be a price equilibrium measure. Then, because  $\frac{S_t}{(1+r)^t}$  is a martingale with respect to  $\nu$ ,  $B_t = (1+r)^t$  and the trading strategy is self-financing, we get

$$\begin{aligned}
& \frac{1}{(1+r)^t} \mathbb{E}_\nu(C_t | \mathcal{F}_{t-1}) \\
&= \frac{1}{(1+r)^t} (\mathbb{E}_\nu(\theta_t S_t | \mathcal{F}_{t-1}) + \mathbb{E}_\nu(\phi_t B_t | \mathcal{F}_{t-1})) \\
&= \theta_t \frac{S_{t-1}}{(1+r)^{t-1}} + \phi_t \frac{B_t}{(1+r)^t} \\
&= \theta_t \frac{S_{t-1}}{(1+r)^{t-1}} + \phi_t \frac{B_{t-1}}{(1+r)^{t-1}} - \frac{1}{(1+r)^{t-1}} ((\theta_t - \theta_{t-1}) S_{t-1} + (\phi_t - \phi_{t-1}) B_{t-1}) \\
&= \theta_{t-1} \frac{S_{t-1}}{(1+r)^{t-1}} + \phi_{t-1} \frac{B_{t-1}}{(1+r)^{t-1}} \\
&= \frac{C_{t-1}}{(1+r)^{t-1}}
\end{aligned}$$

Using this we can compute the price of the portfolio at time  $t = 0$  simply as  $\frac{1}{(1+r)^T} \mathbb{E}_\nu(C_T)$ . This principle is very important in pricing of derivatives which can be replicated by a trading strategy and where the terminal pay-out is known e.g. a European call option on the stock.

Consider the likelihood ratio process  $Z_t = \mathbb{E}_\mu\left(\frac{d\nu}{d\mu} | \mathcal{F}_t\right)$  and put

$$V_t = W_t - \int_0^t \frac{d\langle W, Z \rangle}{Z}$$

By Girsanov's theorem  $V$  is a  $\nu$ -martingale. We get

$$W_t - W_{t-1} = V_t - V_{t-1} + \frac{\langle W, Z \rangle_t - \langle W, Z \rangle_{t-1}}{Z_{t-1}}$$

We can write

$$\begin{aligned}
\frac{S_t}{(1+r)^t} &= S_0 \prod_{s=1}^t \frac{1 + \mu_s + W_s - W_{s-1}}{1+r} \\
&= S_0 \prod_{s=1}^t \frac{1 + \mu_s + V_s - V_{s-1} + \frac{\langle W, Z \rangle_s - \langle W, Z \rangle_{s-1}}{Z_{s-1}}}{1+r}
\end{aligned}$$

Since

$$\frac{S_{t-1}}{(1+r)^{t-1}} = S_0 \prod_{s=1}^{t-1} \frac{1 + \mu_s + V_s - V_{s-1} + \frac{\langle W, Z \rangle_s - \langle W, Z \rangle_{s-1}}{Z_{s-1}}}{1+r}$$

is measurable with respect to  $\mathcal{F}_{t-1}$  we get

$$\begin{aligned} \mathbb{E}_\nu\left(\frac{S_t}{(1+r)^t} \middle| \mathcal{F}_{t-1}\right) &= \mathbb{E}_\nu\left(\frac{S_{t-1}}{(1+r)^{t-1}} \frac{1 + \mu_t + \Delta V_t + \frac{\Delta \langle W, Z \rangle_t}{Z_{t-1}}}{1+r} \middle| \mathcal{F}_{t-1}\right) \\ &= \frac{S_{t-1}}{(1+r)^{t-1}} \mathbb{E}_\nu\left(\frac{1 + \mu_t + \Delta V_t + \frac{\Delta \langle W, Z \rangle_t}{Z_{t-1}}}{1+r} \middle| \mathcal{F}_{t-1}\right) \end{aligned}$$

This shows that  $\nu$  is a price equilibrium measure if and only if

$$\mathbb{E}_\nu\left(\frac{1 + \mu_t + \Delta V_t + \frac{\Delta \langle W, Z \rangle_t}{Z_{t-1}}}{1+r} \middle| \mathcal{F}_{t-1}\right) = 1$$

Since  $V$  is a  $\nu$ -martingale,  $\mathbb{E}_\nu(\Delta V_t | \mathcal{F}_{t-1}) = 0$  and since all the processes in the numerator are predictable this equation is equivalent to

$$1 + \mu_t + \frac{\Delta \langle W, Z \rangle_t}{Z_{t-1}} = 1 + r$$

or

$$\frac{\Delta \langle W, Z \rangle_t}{Z_{t-1}} = -(\mu_t - r)$$

This is the condition the likelihood ratio process has to satisfy in order for  $\nu$  to be a price equilibrium measure.

Put  $\sigma_t^2 = \mathbb{E}_\mu[\Delta W_t^2 | \mathcal{F}_{t-1}] = \Delta \langle W, W \rangle_t$  and put  $\beta_{t-1} = \frac{\mu_t - r}{\sigma_t^2}$ . Remark that  $\beta$  is adapted because  $\mu_t$  and  $\sigma_t^2$  are measurable with respect to  $\mathcal{F}_{t-1}$ . We now define the process which will be a candidate for the likelihood ratio process:

$$Z_t = \prod_{s=1}^t (1 - \beta_{s-1} \Delta W_s) = \mathcal{E}_t(-\beta \bullet W)$$

Remark that  $Z$  is a martingale. Now we have  $\frac{Z_t}{Z_{t-1}} = 1 - \beta_{t-1} \Delta W_t$  or equivalently  $\Delta Z_t = Z_{t-1} \left( -\frac{\mu_t - r}{\sigma_t^2} \right) \Delta W_t$ . Multiplying both sides by  $\Delta W_t$

and taking conditional expectations we get

$$\mathbb{E}_\mu(\Delta W_t \Delta Z_t | \mathcal{F}_{t-1}) = Z_{t-1} \left( -\frac{\mu_t - r}{\sigma_t^2} \right) \mathbb{E}_\mu((\Delta W_t)^2 | \mathcal{F}_{t-1})$$

But  $\mathbb{E}_\mu(\Delta W_t \Delta Z_t | \mathcal{F}_{t-1}) = \Delta \langle W, Z \rangle_t$  and  $\mathbb{E}_\mu((\Delta W_t)^2 | \mathcal{F}_{t-1}) = \sigma_t^2$  so this equation reads  $\Delta \langle W, Z \rangle_t = -Z_{t-1}(\mu_t - r)$  which is precisely the condition for  $Z$  to be the likelihood ratio process of a price equilibrium measure. We can now define  $\nu$  by  $\nu(A) = \int_A Z_T d\mu$ , then  $\nu$  is absolutely continuous with respect to  $\mu$  and  $\sigma$ -additive. The only condition missing for  $\nu$  to be a measure is  $\nu \geq 0$ . Thus we need  $Z_T \geq 0$ . From the definition of  $Z$  this is equivalent to  $1 - \beta_{t-1} \Delta W_t \geq 0$  for all  $t$ . The condition for the existence of a price equilibrium measure is then  $\frac{\mu_t - r}{\sigma_t^2} \Delta W_t \leq 1$  for all  $t$ .

If  $\nu$  is a price equilibrium measure and  $Z_t = \mathbb{E}_\mu\left(\frac{d\nu}{d\mu} | \mathcal{F}_t\right)$  the likelihood ratio process we have by corollary 3.1.1 that  $\left\{Z_t \frac{S_t}{(1+r)^t}\right\}$  is a  $\mu$ -martingale. The process  $Z$  is called the risk adjustment process for this market.

Consider now a market with  $N$  securities, with price processes  $S_1, S_2, \dots, S_N$ . Security 1 is assumed to be a riskless bond so  $S_{1,t} = (1+r)^t$ . Assume there exists a price equilibrium measure  $\nu$  i.e. a measure such that the processes  $\left\{\frac{S_n}{S_1}\right\}_{n=2,\dots,N}$  are martingales with respect to  $\nu$ .

Then we have  $\mu_{n,t} - r = -\frac{\Delta \langle W_n, Z \rangle_t}{Z_{t-1}}$ . We can think of  $\Delta \langle W_n, Z \rangle_t = \text{Covar}(W_{n,t}, Z_t | \mathcal{F}_{t-1})$  as a measure of risk. Thus this equation expresses a linear relationship between expected return and risk.