Introduction to Financial Mathematics

Lecture Notes — MAP 5601

Department of Mathematics
Florida State University

Fall 2003

Table of Contents

1. Finite Probability Spaces

The toss of a coin or the roll of a die results in a finite number of possible outcomes. We represent these outcomes by a set of outcomes called a sample space. For a coin we might denote this sample space by $\{H, T\}$ and for the die $\{1, 2, 3, 4, 5, 6\}$. More generally any convenient symbols may be used to represent outcomes. Along with the sample space we also specify a probability function, or measure, of the likelihood of each outcome. If the coin is a fair coin, then heads and tails are equally likely. If we denote the probability measure by P, then we write $P(H) = P(T) = \frac{1}{2}$. Similarly, if each face of the die is equally likely we may write $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$.

Defninition 1.1. A finite probability space is a pair (Ω, P) where Ω is the sample space set and P is a probability measure:

If $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, then

(i) $0 < P(\omega_i) \leq 1$ for all $i = 1, \ldots, n$

(ii) $\sum_{i=1}^{n} P(\omega_i) = 1$.

In general, given a set of A, we denote the power set of A by $\mathcal{P}(A)$. By definition this is the set of all subsets of A. For example, if $A = \{1, 2\}$, then $\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. Here, as always, $\emptyset$ is the empty set. By additivity, a probability measure on Ω extends to $\mathcal{P}(\Omega)$ if we set $P(\emptyset) = 0$.

Example. For the toss of a fair die, $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$, while P (toss is even) $= P$ (toss is odd) $= \frac{1}{2}$ and $P(\{2, 4, 6\}) = P(2) + P(4) + P(6) = 3 \cdot \frac{1}{6} = \frac{1}{2}$.

The division of $\{1, 2, \ldots, 6\}$ into even and odd, $\{2, 4, 6\}$ and $\{1, 3, 5\}$, is an example of a partition.

Defninition 1.2. A partition of a set Ω (of arbitrary cardinality) is a collection of nonempty disjoint subsets of Ω whose union is Ω.

If the outcome of a die toss is even, then it is an element of $\{2, 4, 6\}$. In this way partitions may provide information about outcomes.

Defninition 1.3. Let $\mathcal{A}$ be a partition of Ω. A partition $\mathcal{B}$ of Ω is a refinement of $\mathcal{A}$ if every member of $\mathcal{B}$ is a subset of a member of $\mathcal{A}$.

For example $\mathcal{B} = \{\{1, 2\}, \{3\}, \{4, 5, 6\}\}$ is a refinement of $\{\{1, 2, 3\}, \{4, 5, 6\}\}$. Notice

that a refinement contains at least as much information as the original partition.

In the language of probability theory, a function on the sample space Ω is called a random variable. This is because the value of such a function depends on the random occurrence of a point of Ω. However, without this interpretation, a random variable is just a function.

Given a finite probability space (Ω, P) and the real-valued random variable $X : \Omega \to \mathbb{R}$ we define the expected value of X or expectation of X to be the weighted probability of its values.

Definition 1.4. The expectation, $E(X)$, of the random variable $X : \Omega \to \mathbb{R}$ is by definition

$$E(X) = \sum_{i=1}^{n} X(\omega_i)P(\omega_i)$$

where $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$.

We see in this definition an immediate utility of the property $\sum_{i=1}^{n} P(\omega_i) = 1$. If X is identically constant, say $X = C$, then $E(X) = C$.

When a partition of Ω is given, giving more informaiton in general than just Ω, we define a conditional expectation.

Definition 1.5. Given a finite probability space (Ω, P) and a partition of Ω, $\mathcal{A}$, we define the conditional expectation of the random variable $X : \Omega \to \mathbb{R}$ with respect to the partition $\mathcal{A}$ at the point $\omega \in \Omega$ by

$$E(X|\,\mathcal{A})(\omega) = \frac{\sum\limits_{\eta \in \mathcal{A}(\omega)} X(\eta)P(\eta)}{P(\mathcal{A}(\omega))}.$$

Here $\mathcal{A}(\omega)$ is the member of $\mathcal{A}$ containing ω and $P(\mathcal{A}(\omega)) = \sum\limits_{\eta \in \mathcal{A}(\omega)} P(\eta)$.

Notice that

$$E(X|\{\Omega\}) = E(X).$$

This holds more generally. When iterating conditional expectations it is the smaller or sparser partition that determines the expected outcomes. A partition $\mathcal{A}$ is smaller than $\mathcal{B}$ if $\mathcal{B}$ is a refinement of $\mathcal{A}$.

Proposition 1.6. *If $\mathcal{B}$ is a refinement of $\mathcal{A}$, then*

$$E(X|\,\mathcal{A}) = E(E(X|\,\mathcal{B})|\,\mathcal{A})$$

$$= E(E(X|\,\mathcal{A})|\,\mathcal{B})$$

for a random variable $X : \Omega \to \mathbb{R}$.

Exercise 1. Prove Proposition 1.6.

Definition 1.7. A random variable $X : \Omega \to \mathbb{R}$ is measurable with respect to a partition $\mathcal{A}$ if X is constant on each set in $\mathcal{A}$.

So, for example, a function constant on Ω is measurable with respect to any partition of Ω. On the other hand if X assumes n distinct values on $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ then X is measurable only with respect to the discrete partition $\{\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_n\}\}$.

Proposition 1.8. *The conditional expectation,* $E(X|\,\mathcal{A})$*, is measurable with respect to* $\mathcal{A}$.

Proof. The proof follows immediately from the definition of $E(X|\,\mathcal{A})$.

Proposition 1.9. *If* X *is measurable with respect to the partition* $\mathcal{A}$*, then* $E(X|\,\mathcal{A}) = X$*. Moreoever, if* Y *is another random variable, then*

$$E(XY|\,\mathcal{A}) = XE(Y|\,\mathcal{A}).$$

Proof. The proof again follows immediately from the definition of $E(X|\,\mathcal{A})$.

We give an interpretation of Proposition 1.9. If X is measurable with respect to $\mathcal{A}$, then the "best guess" of X, given $\mathcal{A}$, is X itself.

Example. Let Ω be the outcomes of a toss of a die and $\mathcal{A} = \{\{2,4,6\}, \{1,3,5\}\}$. Define $X = 1$ if the outcome is even and $X = -1$ otherwise. Then

$$E(X|\,\mathcal{A})(\omega) = \frac{X(\omega)P(\mathcal{A}(\omega))}{P(\mathcal{A}(\omega))} = 1$$

when the outcome is even and -1 otherwise. Notice here the result is the same independent of the probabililties of the individual outcomes. The following properites are immediate consequences of the definition of $E(X|\,\mathcal{A})$.

Proposition 1.10. *Let* X *and* Y *be random variables on* (Ω, P).
(a) If $a, b \in \mathbb{R}$*, then*

$$E(aX + bY|\,\mathcal{A}) = aE(X|\,\mathcal{A}) + bE(Y|\,\mathcal{A}).$$

(b) If $X \leq Y$*, then*

$$E(X|\,\mathcal{A}) \leq E(Y|\,\mathcal{A}).$$

Defninition 1.11. Given a random variable X on Ω, we denote by $p(X)$ the partition generated by X. This is the smallest (coarsest) partition of Ω such that X is measurable with respect to $p(X)$. Notice it follows that X is measurable with respect to a partition $\mathcal{A}$ if and only if $\mathcal{A}$ is a refinement of $p(X)$.

Defninition 1.12. A filtration (of partitions of Ω) is a sequence of partitions

$$\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \ldots, \mathcal{A}_N$$

where

$$\mathcal{A}_{t+1} \text{ is a refinement of } \mathcal{A}_t \text{ for } t = 1, 2, \ldots, N-1.$$

Example. If $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, then $\mathcal{A}_1 = \{\Omega\}$, $\mathcal{A}_2 = \{\{\omega_1, \omega_2\}, \{\omega_3.\omega_4\}\}$. $\mathcal{A}_3 = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\}$ is a filtration of Ω.

A sequence of random variables

$$X_1, X_2, \ldots, X_N \text{ on } \Omega$$

is called a stocastic process or simply a process. We associate a process and a filtration in the following ways.

Defninition 1.13. A sequence of random variables $\{X_t\}_{t=1}^N$ is adapted to the filtration $\{\mathcal{A}_t\}_{t=1}^N$ if X_t is measurable with respect to $\mathcal{A}_t$ for all $t = 1, 2, \ldots, N$.

Example. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. Let $\mathcal{A}_1 = \{\Omega\}$, $\mathcal{A}_2 = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$ and $\mathcal{A}_3 = \{\{\omega_2\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\}$. Then $\{\mathcal{A}_t\}_{t=1}^3$ is a filtration. Let

$$X_1(\omega_i) = 1, \quad i = 1, 2, 3, 4,$$

$$X_2(\omega_i) = \begin{cases} 1, & i = 1, 2 \\ 2, & i = 3, 4 \end{cases},$$

$$X_3(\omega_i) = i, \quad i = 1, 2, 3, 4.$$

The process $\{X_t\}_{t=1}^3$ is adapted to $\{\mathcal{A}_t\}$.

Defninition 1.14. A process $\{X_t\}$ is predictable with respect to $\{\mathcal{A}_t\}$ if X_t is measurable with respect to $\mathcal{A}_{t-1}$ for all t.

We remark that if $\{X_t\}$ is adapted to $\{\mathcal{A}_t\}$, then $\{Y_t\}$, where $Y_t = X_{t-1}$, is predictable with respect to $\{\mathcal{A}_t\}$ (when suitably defined at the initial time).

Defninition 1.15. A process $\{X_t\}$ is a martingale with respect to the filtration $\{\mathcal{A}_t\}$ if

$$E(X_t|\,\mathcal{A}_s) = X_s$$

for $1 \leq s \leq t \leq N$. Notice this condition involves the probability measure P on Ω.

Example. Let X be a random variable on (Ω, P). Define

$$Y_t = E(X|\,\mathcal{A}_t) \text{ for } t = 1, 2, \ldots, N.$$

Then, using the law of iterated expectations, Proposition 1.6, we get

$$\begin{aligned} E(Y_t|\,\mathcal{A}_s) &= E(E(X|\,\mathcal{A}_t)|\,\mathcal{A}_s) \\ &= E(X|\,\mathcal{A}_s) \\ &= Y_s \end{aligned}$$

for $1 \leq s \leq t \leq N$. As such $\{Y_t\}$ is a martingale with respect to $\{\mathcal{A}_t\}$.

We remark that if $\{X_t\}$ is a martingale with respect to $\{\mathcal{A}_t\}$, then $\{X_t\}$ is adapted to $\{A_t\}$. Indeed, if we set $s = t$, then we have

$$E(X_t|\,\mathcal{A}_t) = X_t,$$

which implies that X_t is measurable with respect to $\mathcal{A}_t$.

Proposition 1.16. *A process $\{X_t\}$ is a martingale with respect to $\{\mathcal{A}_t\}$ if and only if*

$$E(X_s|\,\mathcal{A}_{s-1}) = X_{s-1}$$

for all $s = 2, 3, \ldots, N$.

Proof. If $\{X_t\}$ is a martingale, then the conclusion follows from Definition 1.15.

Conversely,

$$\begin{aligned} E(X_t|\,\mathcal{A}_{t-2}) &= E(E(X_t|\,\mathcal{A}_{t-2})|\,\mathcal{A}_{t-1}) \\ &= E(E(X_t|\,\mathcal{A}_{t-1})|\,\mathcal{A}_{t-2}) \end{aligned}$$

by the law of iterated expectations. Hence if $E(X_t|\,\mathcal{A}_{t-1}) = X_{t-1}$, then

$$E(X_t|\,\mathcal{A}_{t-2}) = E(X_{t-1}|\,\mathcal{A}_{t-2})$$

$$= X_{t-2}.$$

Repeating this argument shows that $\{X_t\}$ is a martingale.

Defninition 1.17. Given two processes $\{X_t\}_{t=1}^N$ and $\{Y_t\}_{t=1}^N$, we define the optional quadratic covariation process by

$$[X,Y]_t = \sum_{s=1}^{t} \Delta X_s \Delta Y_s$$

for $1 \le t \le N$. Here $\Delta X_s = X_s - X_{s-1}$ and $X_0 = Y_0 = 0$.

Exercise 2. Prove the polarization identity

$$\boldsymbol{[X,Y]_t = \frac{1}{2}([X+Y, X+Y]_t - [X,X]_t - [Y,Y]_t).}$$

We remark that the so-called optional quadratic variation process of $\{X_t\}$, $\{[X,X]_t\}$ is an increasing process with respect to t.

Defninition 1.18. The predictable quadratic covariation process is defined by

$$\langle X,Y\rangle_t = \sum_{s=1}^{t} E(\Delta X_s \Delta Y_s|\,\mathcal{A}_{s-1})$$

where we also set $\mathcal{A}_0 = \{\Omega\}$.

We remark that since each term is conditioned on the previous partition, $\{\langle X,Y\rangle_t\}$ is a predictable process.

Proposition 1.19. *If $\{X_t\}$ and $\{Y_t\}$ are martingales (for $(\Omega, P, \{\mathcal{A}_t\})$), then so are $\{X_tY_t - [X,Y]_t\}$ and $\{X_tY_t - \langle X,Y\rangle_t\}$.*

Proof. We show that

$$E(X_tY_t - [X,Y]_t|\,\mathcal{A}_{t-1}) = X_{t-1}Y_{t-1} - [X,Y]_{t-1}.$$

It then follows from Proposition 1.16 that $\{X_tY_t - [X,Y]_t\}$ is a martingale. To this purpose notice that

$$[X,Y]_t = \Delta X_t \Delta Y_t + [X,Y]_{t-1}.$$

Furthermore

$$[X,Y]_{t-1} = X_1Y_1 + \Delta X_2\Delta Y_2 + \ldots + \Delta X_{t-1}\Delta Y_{t-1}.$$

Now since martingales are in particular adapted, we have

$$E(\Delta X_s \Delta Y_s |\, \mathcal{A}_{t-1}) = \Delta X_s \Delta Y_s$$

for all $1 \le s \le t-1$. It follows from linearity that

$$E([X,Y]_{t-1} |\, \mathcal{A}_{t-1}) = [X,Y]_{t-1}.$$

We calculate

$$\begin{aligned} E(X_tY_t - [X,Y]_t |\, \mathcal{A}_{t-1}) &= E(X_tY_t - \Delta X_t \Delta Y_t - [X,Y]_{t-1} |\, \mathcal{A}_{t-1}) \\ &= E(X_{t-1}Y_t + X_tY_{t-1} - X_{t-1}Y_{t-1} |\, \mathcal{A}_{t-1}) - [X,Y]_{t-1} \\ &= X_{t-1}E(Y_t |\, \mathcal{A}_{t-1}) + Y_{t-1}E(X_t |\, \mathcal{A}_{t-1}) - X_{t-1}Y_{t-1} - [X,Y]_{t-1}. \end{aligned}$$

Here we have used Proposition 1.9. Now since $\{X_t\}$ and $\{Y_t\}$ are martingales,

$$E(Y_t |\, \mathcal{A}_{t-1}) = Y_{t-1} \text{ and}$$

$$E(X_t |\, \mathcal{A}_{t-1}) = X_{t-1}.$$

The result follows.

Exercise 3. Show that, under the above assumptions, $\{\boldsymbol{X_tY_t - \langle X, Y\rangle_t}\}$ is a martingale.

Defninition 1.20. Two martingales $\{X_t\}, \{Y_t\}$ are orthogonal if $\langle X, Y\rangle_t = 0$ all $1 \le t \le N$.

Theorem 1.21. *Two processes $\{X_t\}$ and $\{Y_t\}$ are orthogonal martingales if and only if $X_1Y_1 = 0$ and $\{X_tY_t\}$ is a martingale.*

Proof. If we have two orthogonal martingales, then $X_1Y_1 = \langle X, Y\rangle_1 = 0$ and Theorem 1.19 shows that $\{X_tY_t\}$ is a martingale.

Conversely, if $\{X_tY_t\}$ is a martingale, then

$$\langle X, Y\rangle_t = X_tY_t - (X_tY_t - \langle X, Y\rangle_t)$$

is also a martingale since the conditions in linear. However by construction $\langle X, Y\rangle_t$ is also predictable. Now any predictable martingale is necessarily constant in t. Here

$$E(\langle X, Y\rangle_s |\, \mathcal{A}_{s-1}) = \langle X, Y\rangle_{s-1},$$

since $\langle X, Y\rangle$ is a martingale, and

$$E(\langle X, Y\rangle_s |\, \mathcal{A}_{s-1}) = \langle X, Y\rangle_s$$

since $\langle X, Y\rangle$ is predictable. Evidently

$$\begin{aligned}\langle X, Y\rangle_N &= \langle X, Y\rangle_{N-1} \\ &= \ldots = \langle X, Y\rangle_1 \\ &= 0\end{aligned}$$

by assumption. This shows that X and Y are orthogonal processes.

We remark that in this proof, for the first time, we have denoted the process $\{X_t\}$ is simply by X. The utility is worth the abbreviation. We can distinguish the process X from a single random variable X by the context.

When an expression involving combinations of random variables is so abbreviated the tacit time is the same. For example

$$X^2 - [X, X] = \{X_t^2 - [X, X]_t\}.$$

Exercise 4. Two martingales $\boldsymbol{X}$ and $\boldsymbol{Y}$ are orthogonal if and only if $\boldsymbol{X_1 Y_1 = 0}$ and

$$\boldsymbol{E(X_t Y_t |\, \mathcal{A}_s) = E(X_t |\, \mathcal{A}_s) E(Y_t |\, \mathcal{A}_s)}$$

for all $\boldsymbol{1 \le s \le t}$.

Example. We now condsider the outcomes of the toss of a fair coin three times. As such the sample space Ω contains $2^3 = 8$ outcomes. Suppose that the toss of a head wins a dollar while an outcome of a tail looses a dollar. Furthermore, let X_t denote the sum of the winnings at time t. The following table lists these quantities.

Ω	1	2	3	X_1	X_2	X_3
ω_1	H	H	H	1	2	3
ω_2	H	H	T	1	2	1
ω_3	H	T	H	1	0	1
ω_4	H	T	T	1	0	-1
ω_5	T	H	H	-1	0	1
ω_6	T	H	T	-1	0	-1
ω_7	T	T	H	-1	-2	-1
ω_8	T	T	T	-1	-2	-3

Next we define a filtration

$$\mathcal{A}_1 = \{\{\omega_1, \omega_2, \omega_3, \omega_4\}, \{\omega_5, \omega_6, \omega_7, \omega_8\}\}$$

$$\mathcal{A}_2 = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_5, \omega_6\}, \{\omega_7, \omega_8\}\},$$

$$\mathcal{A}_3 = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_5\}, \{\omega_6\}, \{\omega_7\}, \{\omega_8\}\}.$$

Notice that $\{X_t\}_{t=1}^3$ is an adapted process with respect to $\{\mathcal{A}_t\}_{t=1}^3$. Moreover, $\{X_t\}$ is a martingale with respect to $\{\mathcal{A}_t\}$.

Exercise 5. Show that X is a martingale.

Exercise 6. Calculate $X^2 - [X, X]$ and $X^2 - \langle X, X\rangle$ and show that they are martingales.

Definition 1.22. For any processes X and Y we define the stochastic sum of Y with respect to X (the discrete stochastic integral) as the process

$$(Y \cdot X)_t = \begin{cases} 0, & \text{if } t = 0 \\ \sum_{s=1}^{t} Y_s \Delta X_s, & \end{cases}$$

when the sum is defined.

Proposition 1.23. *For processes X and Y,*

$$\Delta(Y \cdot X)_t = (Y \cdot X)_t - (Y \cdot X)_{t-1}$$

$$= Y_t \Delta X_t.$$

Proof. The proof is immediate by definition .

Theorem 1.24. *If X is a martingale and Y is a predictable process, then $(Y \cdot X)$ is a martingale.*

Proof. First,

$$E((Y \cdot X)_t | \mathcal{A}_{t-1}) = E(Y_t \Delta X_t | \mathcal{A}_{t-1}) + E((Y \cdot X)_{t-1} | \mathcal{A}_{t-1}).$$

Since X is a martingale and Y is predictable, in particular both are adapted to the given filtration. As such $E((Y \cdot X)_{t-1} | \mathcal{A}_{t-1}) = (Y \cdot X)_{t-1}$. Furthermore, since Y is predictable, Y_t is measurable on $\mathcal{A}_{t-1}$ and so $E(Y_t \Delta X_t | \mathcal{A}_{t-1}) = Y_t E(\Delta X_t | \mathcal{A}_{t-1})$. Finally, since X is a martingale, $E(\Delta X_t | \mathcal{A}_{t-1}) = 0$. All together, $E((Y \cdot X)_t | \mathcal{A}_{t-1}) = (Y \cdot X)_{t-1}$, proving the assertion.

Proposition 1.25. *For processes Y, W and X, we have*

$$(Y \cdot (W \cdot X)) = (YW \cdot X).$$

Proof.

$$\begin{aligned}(Y \cdot (W \cdot X))_t &= \sum_{s=1}^{t} Y_s \Delta (W \cdot X)_s \\ &= \sum_{s=1}^{t} Y_s W_s \Delta X_s\end{aligned}$$

where we have used Proposition 1.23.

Proposition 1.26. *For processes Y, W and X, we have*

$$[(Y \cdot X), W] = (Y \cdot [X, W]).$$

Proof. By definition

$$\begin{aligned}[(Y \cdot X), W]_t &= \sum_{s=1}^{t} \Delta (Y \cdot X)_s \Delta W_s \\ &= \sum_{s=1}^{t} Y_s \Delta X_s \Delta W_s.\end{aligned}$$

On the other hand

$$(Y \cdot [X, W])_t = \sum_{s=1}^{t} Y_s \Delta [X, W]_s.$$

However $\Delta[X,W]_s = [X,W]_s - [X,W]_{s-1} = \Delta X_s \Delta W_s$.

Proposition 1.27. *For processes Y, W and X, assuming that Y is predictable,*

$$\langle (Y \cdot X), W \rangle = (Y \cdot \langle X, W \rangle).$$

Exercise 7. Prove Proposition 1.27.

Proposition 1.28. *Define the process X^- as the right-shift of the process X, that is $X_t^- = X_{t-1}$ with $X_0 = 0$. Then*

$$(X^- \cdot X) = \frac{1}{2}(X^2 - [X,X]).$$

Proof. One calculates

$$2\sum_{s=1}^{t} X_s^- \Delta X_s = \sum_{s=1}^{t} (X_s^2 - X_{s-1}^2) - \sum_{s=1}^{t} (\Delta X_s)^2.$$

The first term on the right telescopes to X_t^2 and the result follows.

Exercise 8. Given two processes $\boldsymbol{X}$ and $\boldsymbol{Y}$, use the polarization identity of Exercise 2 and Proposition 1.28 to show

$$\boldsymbol{(X^- \cdot Y) = XY - (Y^- \cdot X) - [X,Y].}$$

This is the stocastic summation by parts formula.

2. Elements of Continuous Probability Theory

We first define σ-fields.

Definition 2.1. Given an arbitrary set Ω, a collection of subsets of Ω, $\mathcal{F}$ is a σ-field if the following properties hold:

(i) $\phi \in \mathcal{F}$,

(ii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, and

(iii) if $\{A_i\}_{i=1}^{\infty}$ is a countable sequence of sets in $\mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Here $A^c = \Omega \backslash A$ is the complement of A. We remark that $\phi^c = \Omega \in \mathcal{F}$ and so (ϕ, Ω) is the smallest σ-field for any set Ω. On the other hand, notice that the power set, $\mathcal{P}(\Omega)$, is the largest σ-field of Ω.

We remark that σ-fields are used to define integration theory based on measure theory. In probability theory we also use σ-fields to track information as partitions do in the finite case.

Combining properties (ii) and (iii) and using de Morgan's law, we see that, if $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, then

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c \in \mathcal{F}\,.$$

So σ-fields are closed under complementation, countable unions and countable intersections. Notice in the finite case only finite unions and intersections need be considered. In the finite case a σ-field is often called simply a field.

Definition 2.2. When $\mathcal{G}$ is a nonempty collection of subsets of Ω, we write $\sigma(\mathcal{G})$ for the σ-field generated by $\mathcal{G}$. By definition this is the smallest σ-field of Ω containing $\mathcal{G}$.

Exercise 7. Show that $\sigma(\mathcal{G})$ is unique.

Example. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and $\mathcal{G} = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$. Then $\sigma(\mathcal{G}) = \{\phi, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \Omega\}$. If $\mathcal{H} = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\}$, then $\sigma(\mathcal{H}) = \mathcal{P}(\Omega)$.

Notice that the collections $\mathcal{G}$ and $\mathcal{H}$ in the above example are partitions of Ω. In the finite case all σ-fields arise this way.

Theorem 2.3. *Let Ω be finite and let $\mathcal{F}$ be a σ-field of Ω. There is a unique partition $\mathcal{A}$ of Ω so that $\sigma(\mathcal{A}) = \mathcal{F}$.*

Proof. Partially order $\mathcal{F}\setminus\phi$ by inclusion. We claim that $\mathcal{A} = \{\text{minimal elements of } \mathcal{F}\setminus\phi\}$ is a partition of Ω and moreover $\sigma(\mathcal{A}) = \mathcal{F}$. By definition, $A \in \mathcal{F}$ is minimal if there is no proper subset of A in $\mathcal{F}\setminus\phi$. In other words, if $B \in \mathcal{F}\setminus\phi$ and $B \subset A$, then $B = A$. Now if A_1 and A_2 are minimal then $A_1 \cap A_2 = \phi$, otherwise $A_1 = A_2$ or one contains a nonempty proper set $A_1 \cap A_2 \in \mathcal{F}$. Next let $\mathcal{A} = \{A_1, A_2, \ldots, A_k\}$, then $\bigcup_{i=1}^{k} A_i = \Omega$. Otherwise $\left(\bigcup_{i=1}^{k} A_i\right)^c$ is a nonempty set containing a minimal element of $\mathcal{F}\setminus\phi$. Hence $\mathcal{A}$ is a partition of Ω.

Exercise 8. Assume that Ω is finite, $\mathcal{F}$ is a σ-field and $\mathcal{A}$ is the partition of minimal elements of $\mathcal{F}\setminus\phi$ under inclusion. Show that $\sigma(\mathcal{A}) = \mathcal{F}$. Moreover if $\mathcal{F}_1$ and $\mathcal{F}_2$ are two σ-fields of Ω with corresponding partitions $\mathcal{A}_1$ and $\mathcal{A}_2$ and if $\mathcal{F}_1 = \mathcal{F}_2$, then $\mathcal{A}_1 = \mathcal{A}_2$.

So in the finite case there is a one-to-one correspondence between partitions of Ω and σ-fields. However not all σ-fields of infinite sets arise in this way.

An important σ-field of the real numbers is the Borel sets.

Definition 2.4. The Borel subsets of the real line $\mathbb{R}$ is the σ-field generated by the collection of all open sets of $\mathbb{R}$. We denote this σ-field by $\mathcal{B}(\mathbb{R})$.

We remark that since every open subset of $\mathbb{R}$ is a countable union of open intervals, the open intervals also generate $\mathcal{B}(\mathbb{R})$.

Notice that the closed interval $[a, b]$ is a countable intersection of open intervals. Indeed

$$[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right).$$

Hence $[a, b] \in \mathcal{B}(\mathbb{R})$.

Definition 2.5. The pair $(\Omega, \mathcal{F})$, where Ω is a set and $\mathcal{F}$ is a σ-field of subsets is called a measurable space.

Definition 2.6. A measure, μ, defined on the measurable space $(\Omega, \mathcal{F})$, is an extended real-valued set-function defined on sets in $\mathcal{F}$ and satisfying:

(i) $\mu(\phi) = 0$ and

(ii) if $\{E_i\}_{i=1}^{\infty}$ is a disjoint collection of subsets of $\mathcal{F}$, then

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

We call Definition 2.6(ii) countable additivity.

Example. If Ω is finite and P is a probability function on Ω, then, setting $P(\phi) = 0$, $(\Omega, \mathcal{P}(\Omega), P)$ is a measure space.

Proposition 2.7. *If $A \subset B$ are elements of $\mathcal{F}$, then*

$$\mu(A) \leq \mu(B).$$

Proof. Notice $B = A \cup (B \cap A^c)$ is a disjoint union so that

$$\mu(A) = \mu(B) - \mu(B \cap A^c) \leq \mu(B).$$

We now give a brief outline of the construction of Lebesgue measure on the real line $\mathbb{R}$. We wish to define a measure on subsets of $\mathbb{R}$ that extend the natural euclidean length of an interval. To this purpose we define outer measure.

Definition 2.8. For $E \subset \mathbb{R}$, we define the outer measure, $m^\star(E)$, as follows.

$$m^\star(E) = \inf_{E \subset \cup I} \sum_I \ell(I).$$

Here the infimum is over all countable collections of open intervals $\{I\}$ with $E \subset \cup I$.

We have the following properties.

Proposition 2.9.

(i) $m^\star$ *is defined on* $\mathcal{P}(\mathbb{R})$,
(ii) $m^\star(E) \geq 0$ *all* $E \subset \mathbb{R}$.
(iii) if $E_1 \subset E_2$, *then* $m^\star(E_1) \leq m^\star(E_2)$,
(iv) $m^\star(E + y) = m^\star(E)$ *all* $E \subset \mathbb{R}$, $y \in \mathbb{R}$,
(v) $m^\star(A) = 0$ *for any countable set* $A \subset \mathbb{R}$,
(vi) $m^\star(\cup E_i) \leq \sum m^\star(E_i)$ *for any countable collection* $\{E_i\}_{i=1}^\infty$.
Moreover, one can show that
(vii) $m^\star(I) = \ell(I)$ *for any interval* $I \subset \mathbb{R}$.

Exercise 9. Prove Proposition 2.9, (i)–(vi).

It turns out that, even when $\{E_i\}$ is a disjoint collection, that equalith in (vi) may fail. Because of this, $m^\star$ is not a measure on $\mathcal{P}(\mathbb{R})$. However (remarkable) $m^\star$ is a measure on a suitable sub-σ-field of $\mathcal{P}(\mathbb{R})$ (which is in fact very large).

Definition 2.10 (Carathéodory). A set $E \subset \mathbb{R}$ is Lebesgue measurable if

$$m^\star(A) = m^\star(A \cap E) + m^\star(A \cap E^c)$$

for all sets $A \subset \mathbb{R}$.

We write $\mathcal{L}$ for the collection of Lebesgue measurable sets.

Theorem 2.11. *The collection $\mathcal{L}$ is a σ-field of $\mathbb{R}$.*

It is easy to see that ϕ and Ω are in $\mathcal{L}$. From the symmetry in E and Ec, if $E \in \mathcal{L}$, then $Ec \in \mathcal{L}$, follows immediately. What is not so clear, but true, is that $\mathcal{L}$ is closed under countable unions. See [R].

Theorem 2.12. *The Borel subsets, $\mathcal{B}(\mathbb{R})$, are contained in $\mathcal{L}$.*

One only needs to show that an interval is in $\mathcal{L}$. Again, see [R].

Theorem 2.13. *The space $(\mathbb{R}, \mathcal{L}, m)$, where $m = m^\star|_{\mathcal{L}}$, is a measure space.*

In particular, $m^\star$ is countably additive on the σ-field $\mathcal{L}$. See [R].

We now discuss integration theory on a general measure space.

Proposition 2.14. *Let $f : \Omega \to \mathbb{R}$ be a real-valued space $(\Omega, \mathcal{L})$. The following are equivalent:*

(i) $\{x|f(x) > \alpha\} \in \mathcal{F}$ *for all* $\alpha \in \mathbb{R}$,
(ii) $\{x|f(x) \geq \alpha\} \in \mathcal{F}$ *for all* $\alpha \in \mathbb{R}$,
(iii) $\{x|f(x) < \alpha\} \in \mathcal{F}$ *for all* $\alpha \in \mathbb{R}$,
(iv) $\{x|f(x) \leq \alpha\} \in \mathcal{F}$ *for all* $\alpha \in \mathbb{R}$,

Moreover, any of (i)–(iv) implies

(v) $\{x|f(x) = \alpha\} \in \mathcal{F}$ *for all* $\alpha \in \mathbb{R}$,

Definition 2.15. A function $f : \Omega \to \mathbb{R}$, with domain a member of $\mathcal{F}$, is measurable with respect to the σ-field $\mathcal{F}$ if (i)–(iv) holds in Proposition 2.14.

Proof of Proposition 2.14. Assume (i). Then $\{x|f(x) \leq \alpha\} = \bigcap \infty_{n=1} A_n$ where $A_n = \{x|f(x) < \alpha + \frac{1}{n}\}$. This proves (iv). Similarly $\{x|f(x) < \alpha\} = \bigcup \infty_{n=1} \{x|f(x) \leq \alpha - \frac{1}{n}\}$, so that (ii) implies (i). Also, the sets in (iii) and (iv) are complements of thos in (ii) and (i). Proposition 2.14 follows.

INSERT 1 **(NOT YET!)**

Example. Any real-valued function $f : \Omega \to \mathbb{R}$ is measurable with respect to the measurable space $(\Omega, \mathcal{P}(\Omega))$.

Example. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $P(\omega_i) = \frac{1}{4}$ for all $i = 1, 2, 3, 4$ and $\mathcal{A} = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$. We define two random variables on $(\Omega, \sigma(\mathcal{A}), P)$.

Ω	ω_1	ω_2	ω_3	ω_4
X_1	1	1	-1	-1
X_2	1	-1	1	-1

Then X_1 is measurable with respect to $(\Omega, \sigma(\mathcal{A}))$. However X_2 is not. Notice $\{\omega | X_2(\omega) > 0\} = \{\omega_1, \omega_3\}$, which is not an element of $\sigma(\mathcal{A})$.

Proposition 2.1. *A random variable X is measurable with respect to the measurable space $(\Omega, \sigma(\mathcal{A}))$ where Ω is finite and $\mathcal{A}$ is a partition of Ω, if and only if X is measurable with respect to the partition $\mathcal{A}$ (Definition 1.7).*

Example. Let $f(x) = x$. Then $\{x | f(x) > \alpha\}$ is the open set $\{x | x > \alpha\}$. As such $f(x)$ is measurable with respect to $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$.

Definition 2.17. A simple function, φ, on $(\Omega, \mathcal{F})$ is a real-valued function on Ω that assumes a finite number of distinct nonzero values $\{a_1, a_2, \ldots, a_n\}$. Moreover, $A_i = \{\omega | \varphi(\omega) = a_i\}$ is a member of $\mathcal{F}$ for all $i = 1, 2, \ldots, n$.

The representation

$$\varphi(\omega) = \sum_{i=1}^{n} a_i \chi_{\mathcal{A}(\omega)}$$

is called the standard representation of φ. Here $\chi_{\mathcal{A}}(\omega) = 1$ if $\omega \in \mathcal{A}$ and 0 otherwise.

Notice that a sinple function on $\Omega, \mathcal{F})$ is a measurable function on $(\Omega, \mathcal{F})$.

Definition 2.18. When φ is a simple function on the measure space $(\Omega, \mathcal{F}, \mu)$, we define the integral of φ, with respect to μ, as

$$\int \varphi = \int_{\Omega} \varphi d\mu = \sum_{i=1}^{n} a_i \mu(\mathcal{A}_i).$$

Moreover, if $E \in \mathcal{F}$ we define

$$\int_{E} \varphi d\mu = \sum_{i=1}^{n} a_i \mu(\mathcal{A}_i \cap E).$$

Example. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and $P(\omega_i) = \frac{1}{4}$. Define a random variable X on $(\Omega, \mathcal{P}(\Omega), P)$ by $X(\omega_1) = X(\omega_2) = 1$, $X(\omega_3) = X(\omega_4) = -1$. We calculate,

$$\begin{aligned}\int_\Omega X dP &= 1 \cdot P(\{\omega_1, \omega_2\}) + (-1) \cdot P(\{\omega_3, \omega_4\}) \\ &= 1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2} \\ &= 0.\end{aligned}$$

This holds more generally.

Proposition 2.19. *If X is a random variable on a finite probability space $(\Omega, \mathcal{P}(\Omega), P)$, then*

$$\int_\Omega X dP = E(X).$$

Exercise 10. Prove Proposition 2.19.

Definition 2.20. Suppose that f is a nonnegative measurable function on the measure space $(\Omega, \mathcal{F}, \mu)$. We define

$$\int_\Omega f \phi \mu = \sup_{\substack{\varphi \le f \\ \text{simple}}} \int \varphi d\mu.$$

Here the supremum is over all simple functions φ with $\varphi \le f$.

The following theorem allows us to approximate the integral of a nonnegative measurable by a sequence of simple functions increasing to f.

Theorem 2.21 (Monotone Convergence Theorem). *Suppose that $\{f_n\}$ is a sequence of nonnegative measurable functions, $\lim_{n\to\infty} f_n = f$ almost everywhere and $f_n \le f$ for all n. Then*

$$\int_\Omega f d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu.$$

Next, given a nonnegative measurable function f on Ω, we construct simple functions φ_N such that $\lim_{N\to\infty} \varphi_N = f$. Notice that the measurability of f allows us to accomplish this

approximation by partitioning the interval $[0, N)$, $N = 1, 2, 3, \ldots$, into 2^N that converges to zero as $N \to \infty$. Because f is measurable,

$$A_{N,k} = \left\{ \omega | k\frac{N}{2N} \leq f(\omega) < (k+1)\frac{N}{2N} \right\}$$

is in $\mathcal{F}$ for all $N = 1, 2, 3, \ldots$ and all $k = 0, 1, \ldots, 2N - 1$. We define the simple functions φ_N by the first endpoint. That is

$$\varphi_N(\omega) = \sum_{k=1}^{2^N - 1} k\frac{N}{2N}\chi_{A_{N,k}}.$$

By the Monotone Convergence Theorem

$$\lim_{N\to\infty} \int_\Omega \varphi_N d\mu = \int_\Omega f d\mu.$$

Exercise 11. Show that $\lim_{N\to\infty} \varphi_N(\omega) = f(\omega)$. We extend the integral a general $f : \Omega \to \mathbb{R}$ by the decomposition $f = f^+ - f^-$. Here

$$\boldsymbol{F^+ = \max(0, f)} \qquad \text{and} \qquad \boldsymbol{f^- = \max(0, -f).}$$

Exercise 12. Show that if f is measurable, then so is f^+ and f^-.

Next, in preparation for constructing the conditional expectation, we need the following concepts.

Definition 2.22. Given a measure space $(\Omega, \mathcal{F}, \mu)$, we say that a second measure on $(\Omega, \mathcal{F})$, ν, is absolutely continuous with respect to μ if the following condition holods. Whenever $E \in \mathcal{F}$, with $\mu(E) = 0$, we have $\nu(E) = 0$. In other words, sets of measure zero for μ are always sets of measure zero for ν as well.

When ν is absolutely continuous with respect to μ, we write $\nu << \mu$.

Examples.

1. Given $(\Omega, \mathcal{F}, \mu)$ and $D \in \mathbb{C}|\{0\}$, let $\nu = c\mu$. Then $\nu << \mu$ and $\mu << \nu$.
2. Let C be coounting measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. That is $C(E) =$ number of elements in E, when E is finite and ∞ otherwise. If m is the Lebesgue measure on $\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then $n << C$.

3. Let f be a nonnegative measurable function on $(\Omega, \mathcal{F}, \mu)$ and set

$$\nu(E) = \int_E f d\mu$$

for $E \in \mathcal{F}$, then ν is a measure on $(\Omega, \mathcal{F})$ and moreover $\nu << \mu$.

The fact that ν is a measure can be proved from the Monotone Convergence Theorem. The absolute continuity of ν with respect to μ is clear, since integration of f over a set of μ-measure zero equals zero.

It turns out that, in σ-finite cases, the third case is typical. Recall that a measure space $(\Omega, \mathcal{F}, \mu)$ is finite if $\mu(|Omega)$. More generally, if $\Omega = \bigcup \infty_{i=1} \Omega_i$, $\Omega_i \in \mathcal{F}$ and $\mu(\Omega_i)$ for all i, then $(\Omega, \mathcal{F}, \mu)$ is called σ-finite. For example, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$ is a σ-finite measure space since $\mathbb{R} = \bigcup_{N=1}^{\infty} (-N, N)$.

Theorem 2.23 (Radon-Nikodym Theorem). *Suppose that $(\Omega, \mathcal{F}, \mu)$ is a σ-finite measure space and ν is another measure defined on $\mathcal{F}$ with $\nu << \mu$. Then there exists a nonnegative measurable function f such that*

$$\nu(E) = \int_E f d\mu$$

for all $E \in \mathcal{F}$. Moreover if g is any other such function, then $f = g$ a.e. with respect to μ.

We denote the above function f by $\frac{d\nu}{d\mu}$ and refer to it as the Radon-Nikodym derivative of ν with respect to μ. This function has the following properties:

1. If $\nu << \mu$ and f is a nonnegative measurable function, then

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

2. $\dfrac{d(\nu_1 + \nu_2}{d\mu} = \dfrac{d\nu_1}{d\mu} + \dfrac{d\nu_2}{d\mu}$,
3. If $\nu << \mu << \lambda$, then

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda},$$

4. If $\nu << \mu$ and $\mu << \nu$, then

$$\frac{d\nu}{d\mu} = \frac{1}{\frac{d\mu}{d\nu}}.$$

Exercise 13. Prove the above assertions 1–4.

Definition 2.24. A measure space $(\Omega, \mathcal{F}, P)$ is a probability space if $P(\Omega) = 1$.

As such, probability spaces are finite measure spaces.

Examples.

1. Ω finite, P probability function and $(\Omega, \mathcal{P}(\Omega), P)$.
2. $([0,1], \mathcal{L}, m)$.
3. $(\mathbb{R}, \mathcal{L}, f(x)m)$ where f is nonnegative and measurable with $\int_{\mathbb{R}} f dm = 1$.

Definition 2.25. Given a probability space $(\Omega, \mathcal{F}, P)$ and a random variable X we define the expectation of X as

$$E(X) = \int_{\Omega} X dP.$$

Recall that this reduces to Definition 1.4 when Ω is finite and $\mathcal{F} = \mathcal{P}(\Omega)$.

Definition 2.26. We are given a measure space $(\Omega, \mathcal{F}, P)$, a random variable X and a sub-σ-field $\mathcal{G} \subset \mathcal{F}$. The conditional expectation of X with respect to $\mathcal{G}$ is a random variable that we denote by $E(X|\,\mathcal{G})$. It has the following properties:

1. $E(X|\,\mathcal{G})$ is measurable with respect to $\mathcal{G}$,
2. $\int_G X dP = \int_G E(X|\,\mathcal{G}) dP$ for all $G \in \mathcal{G}$.

$E(X|\,\mathcal{G})$ is unique up to sets of measure zero.

3. Differential Equations

Example. Assume that money deposited in a bank increases with continuously compounded interest. This means that the rate of growth is proportional to the amount present. Let $M(t) =$ amount of money present at time t and let r be the proportionality constant. Then r is the constant interest rate and $M(t)$ satisfies the differential equation

$$\frac{dM}{dt} = rM.$$

To solve this equation we relate the differentials using the chain rule.

$$\begin{aligned} dM &= \frac{dM}{dt}\,dt \\ &= rM\,dt. \end{aligned}$$

In this case the variables M and t can be isolated on separate sides of the equation and the equation can be integrated,

$$\int \frac{dM}{M} = \int r\,dt.$$

Or

$$\ln M = rt + C.$$

As such

$$M = Ae^{rt}$$

where $A = e^C$. As such $M(t)$ grows exponentially. Notice the initial amount $M(0) = A$.

Consider the question: How much must be invested today to achieve the amount E at the future time T?

To answer this, one sets the amount, M evaluated at time T equal to the desired future value.

$$E = Ae^{rT}.$$

Solving for A,

$$A = Ee^{-rT}.$$

This is referred to as discounting the future value E.

Example. Let us denote by S the price or value of an asset at time t. For a given t_1 and t_2 we write the change in value by $\Delta S = S(t_2) - S(t_1)$ over the time interval $\Delta t = t_2 - t_1$. The

relative price change, or return, compares the absolute price change to the initial value. If the relative price change is proportional to the time interval, then

$$\frac{\Delta S}{S} = \mu \Delta t.$$

Here μ is a constant called the drift. We can rewrite this as

$$\frac{\Delta S}{\Delta t} = \mu S.$$

Letting $\Delta t \to 0$ we obtain an equation of the same form. Namely,

$$\frac{dS}{dt} = \mu S.$$

These equations are examples of a special type of first-order equations.

Definition 3.1. We here consider y as a function of x. Equations of the form

$$\frac{dy}{dx} = f(x)g(y)$$

are called separable. Here $f(x)$ is a function only of x and $g(y)$ is a function only of y. The variables can be separated and integrated (provided the integrals exist) as indicated.

$$\int \frac{dy}{g(y)} = \int f(x)\, dx.$$

Another often occurring type of first-order differential equation is the linear equation. In these equations y and y' are present, however no other functions of them occur. We may express such an equation in the form

$$y' + p(x)y = q(x).$$

Here $p(x)$ and $q(x)$ are arbitrary functions of x. We pause to remark that these equations are separable when $q(x) \equiv 0$. We call these equations homogeneous. Indeed in this case

$$\int \frac{dy}{y} = -\int p(x)\, dx.$$

Otherwise we solve these equations by multiplying by a so-called integrating factor. Let

$$I(x) = e^{\int^x p(t)\, dt}.$$

We remark that integrating factors are in general not unique. In particular, here any multiple of $I(x)$ will work. As such the lower limit in the definite integral $\int^x p(t)\,dt$ is irrelevant and may be presupposed conveniently chosen. Multiplying by $I(x)$ gives $I(x)y' + I(x)p(x)u = I(x)q(x)$. The key to this procedure is that the left-hand side is the derivative of the product of the solution y and the integrating factor $I(x)$. Indeed

$$\frac{d}{dx}(I(x)y) = I(x)y' + y\frac{d}{dx}I(x).$$

However

$$\begin{aligned}\frac{d}{dx}I(x) &= \frac{d}{dx}\exp\left(\int^x p(t)\,dt\right)\\ &= p(x)\exp\left(\int^x p(t)\,dt\right)\end{aligned}$$

by the chain rule. As such

$$\frac{d}{dx}(I(x)y) = I(x)y' + p(x)I(x)y.$$

Integrating with respect to x gives

$$I(x)u = \int^x I(t)q(t)\,dt$$

and dividing by $I(x)$ obtains the solution.

Example. Let

$$y' + \frac{1}{x}y = x^2.$$

Then

$$\begin{aligned}I(x) &= \exp\left(\int^x \frac{1}{t}\,dt\right)\\ &= \exp(\ln x)\\ &= x.\end{aligned}$$

We have actually made a choice of multiple here. Multiplying by $I(x)$ gives

$$xy' + y = x^3.$$

So

$$\frac{d}{dx}(I(x)y) = \frac{d}{dx}(xy)$$

$$= x^3$$

and

$$xy = \frac{x^4}{4} + C.$$

This gives

$$y = \frac{x^3}{4} + \frac{C}{x}.$$

Example. Suppose the price of an asset, S, grows at a constant rate C. Then

$$\frac{dS}{dt} = C.$$

This equation is separable and

$$S = Ct + C_1.$$

Now let's suppose that there are two contributions to the growth of S. The first is proportional to the asset price and the second a constant rate C. Then

$$\frac{dS}{dt} = \mu S + C,$$

or

$$\frac{ds}{dt} - \mu S = C.$$

This is a linear equation and

$$I(t) = \exp\left(\int^x -\mu\, dS\right)$$

$$= e^{-\mu t}.$$

So

$$\frac{d}{dt}\left(e^{-\mu t}S\right) = Ce^{-\mu t}$$

and

$$S = e^{\mu t}\left[-\frac{C}{\mu}e^{-\mu t} + C_1\right]$$

$$= -\frac{c}{\mu} + C_1 e^{\mu t}.$$

Now the initial amount is $S(0) = C_1 = C/\mu$, so that C_1 is larger than $S(0)$. Notice that $S(t)$ asymtotically approaches $C_1 e^{\mu t}$ as $t \to \infty$.

We now consider solutions to equations of the form

$$y'' + p(x)y' + q(x)y = r(x).$$

These equations are second order, meaning that y'' is the highest derivative that is present. They are linear, meaning that the solution y and its derivatives y' and y'' occur only to first power. Otherwise, $p(x)$, $q(x)$ and $r(x)$ are arbitrary functions of x.

We also consider the so-called associated homogeneous equation:

$$y'' + p(x)y' + q(x)y = 0.$$

Theorem 3.2.

(i) *If y_1 and y_2 are two solutions to the homogeneous equation, then so is their general linear combination, namely*

$$y = C_1 y_1 + C_2 y_2$$

REVISION_____STOPS_____HERE

for arbitrary constants C_1 and C_2.

(ii) *If y is a solution to (2.1) and if y_h is a solution to (2.2), then the sum $y + y_h$ is a solution to (2.1).*

(iii) *If u and v are solutions to (2.1), then the difference is a solution to (2.2).*

Proof.

(i)
$$\begin{aligned}(c_1y_1 + c_2y_2)'' + p(x)(c_1y_1 + c_2y_2)' + q(x)(c_1y_1 + c_2y_2) \\ = c_1[y_1'' + p(x)y_1' + q(x)y_1] + c_2[y_2'' + p(x)y_2' + q(x)y_2] \\ = 0.\end{aligned}$$

(ii)
$$\begin{aligned}(y + y_h)'' + p(x)(y + y_h)' + q(x)(y + y_h) \\ = [y'' + p(x)y' + q(x)y] + [y_h'' + p(x)y_h' + q(x)y_h] \\ = r(x) + 0 \\ = r(x).\end{aligned}$$

(iii)
$$\begin{aligned}(u-v)'' + p(x)(u-v)' + q(x)(u-v) &= [u'' + p(x)u' + q(x)u] - [v'' + p(x)v' + q(x)v] \\ &= r(x) - r(x) \\ &= 0.\end{aligned}$$

Notice that it follows from Theorem 2.3(iii) that the most general solution to (2.1) is ***any*** solution to (2.1) plus the general homogeneous solution. □

We now consider solutions to homogeneous equations with constant real coefficients:

$$y'' + ay' + by = 0. \tag{2.4}$$

The following technique works for such equations of any order. However, here we restrict our attention to second-order equations and, because of the homogeneity, may assume that the coefficient of y'' is one.

One can prove that there are two solutions to (2.4) which are linearly independent. Two functions are linearly independent when they are not multipliers of each other or, what is the same, their ratio is not constant.

Solutions to (2.4) are linear combinations of solutions of the form $y = e^{\lambda x}$ where λ is, in general, a complex constant. As such, every solution (2.4) is a linear combination of the two linearly independent solutions. Indeed, upon substitution into (2.4) one obtains

$$\lambda^2 e^{\lambda x} + a\lambda e^{\lambda x} + be^{\lambda x} = 0,$$

which can only be satisfied if

$$\lambda^2 + a\lambda + b = 0. \tag{2.5}$$

(In fact $e^{\lambda x} \neq 0$ for any λ or any x.) The solutions to (2.5) are given by

$$\lambda = \frac{-a \pm \sqrt{a^2 - 4b}}{2}.$$

Of course in some cases (2.5) can be factored by inspection. Since the coefficients, a and b are assumed real, the solutions to (2.5) and as such to (2.4) fall into three cases.

Case 1. The roots to (2.5) are real and distinct, say λ_1 and λ_2. In this case $e^{\lambda_1 x}$ and $e^{\lambda_2 x}$ are solutions. Since (2.4) is linear, linear combinations of solutions are again solutions. So

$$y = c_1 e^{\lambda_1 x} + c_2 e^{\lambda_2 x}$$

is a solution for any constants c_1 and c_2.

Case 2. The solutions to (2.5) are real and repeated. This happens when the left-hand side of (2.5) is a perfect square. If the solution is λ, then $e^{\lambda x}$ is a solution. The other, linearly independent solution to the first, is $xe^{\lambda x}$. This can be derived from a reduction of otder argument. As such

$$y = c_1 e^{\lambda x} + c_2 x e^{\lambda x}$$

is the general solution. We warn you that a solution of the form $xe^{\lambda x}$ exists only when λ is a repeated root.

Case 3. The solutions are complex numbers. All complex numbers can be expressed in the form $a + ib$ where a and b are real numbers and $i^2 = -1$. When λ_1 and λ_2 are complex solutions to (2.5) they necessarily are complex conjugates. By this we mean if $\lambda_1 = a + ib$, then $\lambda_2 = a - ib$. One can see this from the quadratic formula since, with real coefficients, $(-a/2) \pm (1/2)\sqrt{a^2 - 4b}$ are complex conjugate numbers when $a^2 - 4b < 0$. One can also see this by conjugating equation (2.5). Often the complex conjugate of $\lambda = a + ib$ is denoted $\overline{\lambda} = a - ib$. It is easy to see that conjugation distributes over addition and multiplication of complex numbers. As such $\overline{\lambda^2 + a\lambda + b} = \overline{0}$ or $(\overline{\lambda})^2 + a\,\overline{\lambda} + b = 0$. This show that if λ is a solution, then so is $\overline{\lambda}$. It follows that $e^{(a+ib)x}$ and $e^{(a-ib)x}$ are solutions to (2.4). It turns out that general linear combinations of the complex exponentials can be rewritten equivalently as general linear combinations of functions which are real-valued. Euler's equation expresses complex exponentials as complex linear combinations of the sinusodial functions. Namely, for any real number θ,

$$e^{i\theta} = \cos\theta + i\sin\theta. \tag{2.6}$$

As such

$$e^{ibx} = \cos bx + i\sin bx \tag{2.7}$$

and, because the cosine is even and the sine is odd,

$$e^{-ibx} = \cos bx - i\sin bx. \tag{2.8}$$

Using (2.7) and (2.8), general linear combinations of the complex exponentials can be rewritten as linear combinations of sine and cosine. (Try it!) In the general solution of equation (2.4), the real exponentials from the term a remain as factors. As such solutions have the form

$$y = c_1 e^{ax} \cos bx + c_2 e^{ax} \sin bx.$$

Example 2.9. The equation

$$y'' - 5y' + 6y = 0$$

gives rise to the characteristic equation

$$\lambda^2 - 5\lambda + 6 = 0.$$

The roots are $\lambda = -2$ and -3. As such

$$y = c_1 e^{2x} + c_2 e^{3x}$$

is the general solution.

Example 2.10. Case 2.

$$y'' + 6y' + 9y = 0.$$

Here $(\lambda + 3)^2 = 0$ and $\lambda = -3$ is a double real root. As such the general solution is

$$y = c_1 e^{-3x} + c_2 x e^{-3x}.$$

Example 2.11. As an example of Case 3 we solve

$$y'' + y = 0.$$

Now $\lambda^2 + 1 = 0$ and $\lambda = \pm i$. We can express the general result in terms of real-valued functions

$$y = c_1 \cos x + c_2 \sin x.$$

Suppose we compare the solutions to

$$y'' + y' + y = 0.$$

In this case $\lambda^2 + \lambda + 1 = 0$ and $\lambda = -\frac{1}{2} \pm i\frac{\sqrt{3}}{2}$. The solution now has real exponentials multiplying the sinusoidial solutions

$$y = c_1 e^{-\frac{x}{2}} \cos \frac{\sqrt{3}}{2} x + c_2 e^{-\frac{x}{2}} \sin \frac{\sqrt{3}}{2} x.$$

3. The Dirac Delta Function

The Dirac delta function is a linear mapping which associates a real number to each real-valued continuous function. Such a mapping is called a linear functional. If $\mathbb{R}$ denotes the real line and $C(\mathbb{R})$ the continuous functions defined on $\mathbb{R}$, then the δ-function maps $C(\mathbb{R})$ onto $\mathbb{R}$. Given $\varphi \in C(\mathbb{R})$ we define

$$\delta(\varphi) = \varphi(0).$$

That is δ maps each continuous function to its value at the origin.

We can construct the action of δ by integrating a continuous function φ against a family of functions $\delta_\varepsilon(x)$ and taking a limit. This gives rise to the view of δ as a "function". Translations of the δ function arise naturally in this formalism and the continuous functions are evaluated in this way at all real values. To this purpose we define

$$\delta_\varepsilon(x) = \begin{cases} \frac{1}{2\varepsilon}, & \text{when } -\varepsilon < x < \varepsilon \\ 0, & \text{otherwise} \end{cases}$$

for $\varepsilon > 0$. Notice

$$\int_{-\infty}^{\infty} \delta_\varepsilon \, dx = 1$$

for all ε.

Theorem 3.1. *Let $\varphi \in C(\mathbb{R})$. Then*

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \delta_\varepsilon(x)\varphi(x) \, dx = \varphi(0). \tag{3.2}$$

Proof. Notice

$$\int_{-\infty}^{\infty} \delta_\varepsilon(x)\varphi(x) \, dx = \int_{-\varepsilon}^{\varepsilon} \frac{1}{2\varepsilon}\varphi(x) \, dx.$$

On the other hand

$$\varphi(0) = \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} \varphi(0) \, dx.$$

Hence

$$\left| \int_{-\infty}^{\infty} \delta_\varepsilon(x)\varphi(x)\,dx - \varphi(0) \right| = \frac{1}{2\varepsilon} \left| \int_{-\infty}^{\infty} (\varphi(x) - \varphi(0))\,dx \right|$$

$$\leq \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} |\varphi(x) - \varphi(0)|\,dx.$$

Since φ is continuous at 0, given $\eta > 0$, there exists $\varepsilon > 0$ such that

$$|\varphi(x) - \varphi(0)| < \eta$$

whenever $|x| < \varepsilon$. This proves (3.2). □

It is useful in calculations to write $\delta(x)$ in the integrand and so

$$\int_{-\infty}^{\infty} \delta(x)\varphi(x)\,dx = \varphi(0).$$

Notice if $a < 0$ and $b > 0$,

$$\int_{a}^{b} \delta(x)\,dx = 1$$

in particular, and

$$\int_{a}^{b} \delta(x)\varphi(x)\,dx = \varphi(0).$$

Notice that it follows that

$$\int_{-\infty}^{\infty} \delta(a - x)\varphi(x)\,dx = \varphi(a)$$

for all $a \in \mathbb{R}$. Equivalently

$$\int_{-\infty}^{\infty} \delta(x - a)\varphi(x)\,dx = \varphi(a).$$

and one may think of $\delta(x - a)$ as translating the concentration of the delta function from the origin to the point a.

Example 3.3. Suppose at time $t = 0$ the value of an asset is $V(0)$ and this value increases instantaneously at $t = 1$ by an amount A. The differential equation

$$\frac{dV}{dt} = A\delta(t-1) \tag{3.4}$$

describes this. Integrating (3.4) with respect to t gives

$$\begin{aligned} V(t) - V(0) &= \int_0^t \frac{dV}{d\tau}\, d\tau. \\ &= \int_0^t A\delta(\tau - 1)\, d\tau \\ &= \begin{cases} 0, & 0 < t < 1 \\ A, & t > 1. \end{cases} \end{aligned}$$

If we define the unit step function

$$u(x-a) = \begin{cases} 0, & x < a \\ 1, & x > a, \end{cases}$$

then we can write

$$V(t) = Au(x-a) + V(0).$$

In this sense we symbolically write

$$\frac{d}{dt}u(t-a) = \delta(t-a).$$

4. The Laplace Transform

Given a function $f(t)$, for $t > 0$, we define its Laplace transform

$$\mathcal{L}(f)(s) = \int_0^\infty f(t)e^{-st}\,dt. \tag{4.1}$$

Notice that the improper integral (4.1) is defined for a wide range of functions $f(t)$ because when $s > 0$, e^{-st} is a decaying exponential. In most applications of the Laplace transform which we will encounter, restricting the domain of the variable s will not cause difficulty.

We offer the following table and discuss some of the calculations. Here, and throughout this section, we denote the Laplace transform of $f(t)$ by $F(s)$. When the context permits, we will also represent the Laplace transform of functions denoted by lower case roman letters, by their capitals.

4.2. A Table of transforms. Here the constant a is, in general, a complex nunmber.

	$f(t)$	$F(s)$
a)	t^n	$\frac{n!}{s^{(n+1)}}$
b)	e^{at}	$\frac{1}{s-a}$
c)	$\cos at$	$\frac{s}{s^2+a^2}$
d)	$\sin at$	$\frac{a}{s^2+a^2}$
e)	$\cosh at$	$\frac{s}{s^2-a^2}$
f)	$\sinh at$	$\frac{a}{s^2-a^2}$
g)	$u(t-a)$	$\frac{e^{-as}}{s}$

With some practice one remembers this table without having to recompute. We emphasize that the Laplace transform is linear. As such

$$\mathcal{L}(af + bg) = a\,\mathcal{L}(f) + b\,\mathcal{L}(g),$$

for functions $f(t)$, $g(t)$ and constants a and b. The functions 4.2 c), d), e) and f) are linear combinations of functions of the form e^{at}. Because the Laplace transform is linear, the

transforms of c), d), e) and f) are linear combinations of $\frac{1}{s-a}$. This helps remember the pattern of transforms in Table 4.2.

The Laplace transform transforms linear ordinary differential equations into algebraic equations of the transforms. This is the key property that we here exploit.

Theorem 4.3. *So that the Laplace transforms exist, we assume that there exists an $s > 0$ such that $\lim_{t\to\infty} |f(t)e^{-st}| = 0$. Also we assume that $f'(t)$ exists except perhaps for a finite number of points in any bounded interval. Then*

$$\mathcal{L}(f')(s) = s\,\mathcal{L}(f)(s) - f(0). \tag{4.4}$$

Under appropriate conditions one can iterate (4.4) to get

$$\begin{aligned} \mathcal{L}(f'')(s) &= s\,\mathcal{L}(f')(s) - f'(0) \\ &= s^2\,\mathcal{L}(f)(s) - sf(0) - f'(0). \end{aligned} \tag{4.4}$$

Notice that the initial conditions $f(0)$ and $f'(0)$ are in these formulas. As such solutions to differential equations using the Laplace transform satisfy prescribed initial conditions which gives values for the integration constants in the general solutions. This aspect of the Laplace transform is often convenient.

Example 4.6. Solve the differential equation

$$y'' - 5y' + 6y = 0$$

with initial conditions $y(0) = 1$ and $y'(0) = 0$. Using (4.4), (4.5) and the linearity of $\mathcal{L}$ we get

$$\mathcal{L}(y'') - 5\,\mathcal{L}(y') + 6\,\mathcal{L}(y) = \mathcal{L}(0)$$

5. The Fourier Transform

6. The Diffusion Equation

The one-dimensional diffusion equation (or heat equation) is a partial differential in two variables x and t. In most applications of the diffusion equation the solution $u(x,t)$ is a function in the "spatial" variable x and the variable of time, t. The diffusion equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \tag{6.1}$$

is a parabolic partial differential equation.

Problem 6.2. Solve (6.1) over the real line $-\infty < x < \infty$ for $u = u(x,t)$ with prescribed initial condition $u(x,0) = u_0(x)$.

When the integral exists, the solution to Problem 6.2 is given by

$$u(x,t) = \frac{1}{2\sqrt{\pi t}} \int_{-\infty}^{\infty} u_0(v) e^{-\frac{(x-v)^2}{4t}}\, dv. \tag{6.3}$$

The so-called fundamental solution to Problem 6.2 is the above kernel

$$u_\delta(x,t) = \frac{1}{2\sqrt{\pi t}} e^{-\frac{x^2}{4t}}. \tag{6.4}$$

In this case

$$\lim_{t\to 0} u_\delta(x,t) = \delta(x)$$

in the distributional sense (3.2). As such from (6.3),

$$\begin{aligned}
\lim_{t\to 0} u(x,t) &= \lim_{t\to 0} \frac{1}{2\sqrt{\pi t}} \int_{-\infty}^{\infty} u_0(v) e^{-\frac{(x-v)^2}{4t}}\, dv \\
&= \int_{-\infty}^{\infty} u_0(v)\delta(x-v)\, dv \\
&= u_0(x).
\end{aligned}$$

We give two derivations of (6.3). Because the domain is the entire real line we can directly apply the Fourier transform.

1. Applying the Fourier transform to (6.1) we get

$$\frac{\partial}{\partial t}\,\widehat{u}(s,t) = 4\pi^2 s^2\,\widehat{u}(s,t). \tag{6.5}$$

Solving (6.5),

$$\widehat{u}(s,t) = Ce^{-4\pi^2 s^2 t}$$

and $\widehat{u}(s,0) = C$. On the other hand

$$\widehat{u}_\delta(s,t) = e^{-4\pi^2 s^2 t}, \tag{6.6}$$

Hence

$$\widehat{u}(s,t) = \widehat{u}_0(s)\,\widehat{u}_\delta(s,t). \tag{6.7}$$

The convolution theorem ( ) gives

$$u(x,t) = \int_{-\infty}^{\infty} u_0(v)u_\delta(x-v,t)\,dv$$

$$= (u_0 \star u_\delta)(x,t).$$

This is (6.3).

2. It is useful to realize that (6.3) can also be derived from the superposition of so-called separated solutions. This technique is prevalent when building solutions to linear equations. To this purpose we set $u(x,t) = F(x)G(t)$. Then substitution into (6.1) gives $F(x)G'(t) = F''(x)G(t)$ or

$$\frac{F(x)}{F''(x)} = \frac{G(t)}{G'(t)}. \tag{6.8}$$

Since the left- and right-hand sides of (6.8) depend separately on the independent variables x and t, they must be constant. We choose this constant to be $\frac{1}{-p^2}$. The solutions in this case will decay as $t \to \infty$. As such

$$F''(x) + p^2 F(x) = 0 \tag{6.9}$$

and

$$G'(t) + p^2 G(t) = 0. \tag{6.10}$$

Solving these equations one gets in general

$$F(x) = A\cos px + B\sin px$$

and

$$G(t) = Ce^{-p^2t}.$$

As such, for general A and B,

$$u(x,t) = (A\cos px + B\sin px)e^{-p^2t}$$

is a solution to (6.1).

Now since (6.1) is linear, a superposition of solutions is also a solution. The integral is a limit of Riemann sums which are finite superpositions of solutions. Here we can pass to the limit obtaining

$$U(x,t) = \int_0^\infty (A(p)\cos px + B(p)\sin px)e^{-p^2t}\,dp. \tag{6.11}$$

as a solution to (6.1). Notice that (6.11) is a Fourier integral. In particular

$$U(x,0) = \int_0^\infty (A(p)\cos px + B(p)\sin px)\,dp$$

will equal $u_0(x)$ if we choose

$$A(p) = \frac{1}{\pi}\int_{-\infty}^\infty u_0(v)\cos pv\,dv$$

and

$$B(p) = \frac{1}{\pi}\int_{-\infty}^\infty u_0(v)\sin pv\,dv.$$

Inserting the formulas for the coefficients $A(p)$ and $B(p)$ into (6.11) we obtain

$$\int_0^\infty \left[\frac{1}{\pi}\int_{-\infty}^\infty u_0(v)\cos pv\,dv\ \cos px + \frac{1}{\pi}\int_{-\infty}^\infty u_0(v)\sin pv\,dv\ \sin px\right] e^{-p^2t}\,dp$$

$$= \frac{1}{\pi}\int_0^\infty\int_{-\infty}^\infty u_0(v)[\cos pv\cos px + \sin pv\sin px]e^{-p^2t}\,dv\,dp.$$

Using the summation formula for the cosine (an equation worth remembering, $\cos(A+B) = \cos A \cos B - \sin A \sin B$), we collapse the above integral into

$$\frac{1}{\pi} \int_0^\infty \int_{-\infty}^\infty u_0(v) \cos(pv - px) e^{-p^2 t} \, dv \, dp. \tag{6.12}$$

We next interchange the order of integration and change variables as follows:

$$s = \sqrt{t}\, p \quad \text{and} \quad 2b\sqrt{t} = x - v.$$

Then (6.12) becomes

$$\frac{1}{\pi\sqrt{t}} \int_{-\infty}^\infty u_0(v) \int_0^\infty e^{-s^2} \cos 2bs \, ds \, dv. \tag{6.13}$$

Since

$$\int_0^\infty e^{-s^2} \cos 2bs \, ds = \frac{\sqrt{\pi}}{2} e^{-b^2},$$

we again obtain (6.3).

We next show that certain equations of interest can be solved by solving the diffusion equation.

Proposition 6.14. *Solutions to*

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} + a \frac{\partial v}{\partial x} + bv, \tag{6.15}$$

where a and b are constants, are of the form

$$v(x,t) = e^{[-\frac{a}{2}x + (b - \frac{a^2}{4})t]} u(x,t). \tag{6.16}$$

Here $u(x,t)$ solves the diffusion equation (6.1).

Proof. If v has the form (6.16), then substitution into (6.15) shows that $u(x,t)$ solves (6.1).

Conversely, if v is a solution to (6.15), then $e^{\frac{a}{2}x + (\frac{a^2}{4} - b)t} v$ solves (6.1).

Proposition 6.17. *For $w = w(X,t)$ suppose that, with $A \neq 0$,*

$$\frac{\partial w}{\partial t} = AX^2 \frac{\partial^2 w}{\partial X^2} + BX \frac{\partial w}{\partial X} + Cw. \tag{6.18}$$

There are solutions to (6.18) of the form $v(x,t) = w(X(x),t)$ where $X(x) = e^x$ and v satisfies (6.15) with

$$a = (B - A)/A$$

and

$$b = C.$$

7. The Black-Scholes Equation

8. An Introduction to Measure and Integration

The concept of measure gives a unifying theory for many mathematical situations which may at first seem quite different. The discussion below shows how it arises naturally in the theory of the Lebesgue integral.

Let us first consider the example of a finite probability space. Here X is a set (of outcomes) called a sample space. A probability function P is defined on the set of all subsets of X. (This set is called the power set of X and denoted $\mathcal{P}(X)$.)

By definition a probability function is a function which associates a nonnegative real number to a given subset and which satisfies two conditions:

$$P(X) = 1 \tag{8.1}$$

$$P(A \cup B) = P(A) + P(B) \tag{8.2}$$

when $A \cap B = \emptyset$; i.e., A and B are disjoint. Notice that these conditions imply that $P(A) \leq 1$ for any subset $A \in \mathcal{P}(X)$ since $P(X) = P(A) + P(A^c)$ and in particular, $P(\emptyset) = 0$.

Example 8.3. Consider the outcomes of tossing a coin, H for heads, T for tails. The sample space is

$$X = \{H, T\}.$$

The power set of X is

$$\mathcal{P}(X) = \{\emptyset, \{H\}, \{T\}, X\}.$$

If the coin is fair, then

$$P(\{H\}) = P(\{T\}) = \frac{1}{2}$$

and

$$\begin{aligned} P(X) &= P(\{H\} \cup \{T\}) \\ &= P(\{H\}) + P(\{T\}) \\ &= \frac{1}{2} + \frac{1}{2} \\ &= 1 \end{aligned}$$

Notice that the power set $\mathcal{P}(X)$ of any set X, finite or not, trivially satisfies the properties of an algebra of subsets:

Definition 8.4. A set of subsets of a set X, $\mathcal{A}$, is an algebra of subsets of X satisfying two properties

$$\text{If } A \in \mathcal{A}, \text{ then } A^c = X \backslash A \in \mathcal{A}, \tag{8.5}$$

$$\text{If } A, B \in \mathcal{A}, \text{ then } A \cup B \in \mathcal{A}. \tag{8.6}$$

Notice $A \cup A^c = X$ and so $X \in \mathcal{A}$ and so $X^c = \emptyset \in \mathcal{A}$. Also de Morgan's law gives

$$A \cap B = (A^c \cup B^c)^c$$

and so $A \cap B \in \mathcal{A}$. Notice that $(X, \emptyset)$ is the smallest algebra of subsets of X.

In the case of a finite probability space, X, and therefore $\mathcal{P}(X)$, is finite. In other cases X is infinite and it is often necessary to consider sets of subsets of X which are smaller than $\mathcal{P}(X)$ but still retain the properties of an algebra.

Since, when X is infinite, algebras of subsets may also be infinite, it is necessary to extend (8.3) as follows.

Definition 8.7. A σ-algebra, $\mathcal{F}$, of subsets of X is an algebra of subsets such that

$$\text{If } \{A_i\}_{i=1}^{\infty} \text{ is a sequence of sets in } \mathcal{F}, \text{ then their union } \bigcup_{i=1}^{\infty} A_i \text{ is also in } \mathcal{F}. \tag{8.8}$$

Example 8.9. Let $X = \mathbb{R}$ and $\mathcal{A} =$ set of finite unions of half-open intervals. That is, finite unions of sets of the form $[a, b)$, where $-\infty < a \leq b \leq \infty$, along with the intervals $(-\infty, a)$. Then $\mathcal{A}$ is an algebra of subsets of $\mathbb{R}$.

Example 8.10. Given a collection of subsets $\mathcal{C} \subset \mathcal{P}(X)$ there is a smallest σ-algebra containing $\mathcal{C}$. We call this σ-algebra the σ-algebra generated by $\mathcal{C}$ and denote it by $\sigma(\mathcal{C})$.

For $X = \mathbb{R}$ and $\mathcal{A}$ as in Example 8.9, $\sigma(\mathcal{A})$ is called the Borel subsets of $\mathbb{R}$ and denoted by $\mathcal{B}$.

The Borel subsets $\mathcal{B}$ is also the σ-algebra generated by the open sets of $\mathbb{R}$, since each open set is a countable union of open intervals.

Example 8.11. On the other hand consider the collection of intervals of $\mathbb{R}$, $\mathcal{C} = \{(-\infty, 0), [0, \infty)\}$. Notice the intervals partition $\mathbb{R}$, that is they are disjoint and $\mathbb{R}$ is their union. In this case

$$\sigma(\mathcal{C}) = \{\emptyset, (-\infty, 0), [0, \infty), \mathbb{R}\}.$$

Also notice that a similar result occurs when $\mathcal{C}$ is any partition of $\mathbb{R}$.

We next discuss the concept of a measure on sets. A probability function on a probability space is an important example.

We motivate the concepts by outlining the construction of the Lebesgue integral, which requires Lebesgue measure. To this purpose we first review the Riemann integral.

The Riemann integral, when it exists, is defined to be a limit of so-called Riemann sums. We now describe these sums.

Suppose that $f(x)$ is defined on a closed interval $[a, b]$. We partition this interval into subintervals labeled $x_0 = a < x_1 < x_2 < \cdots < x_n = b$. In each subinterval $[x_{i-1}, x_i]$ we choose (sample) a functional value $f(x_i^\star)$ where $x_{i-1} \le x_i^\star \le x_i$. This functional value is used as an approximation to the functional values of $f(x)$ over the subinterval $[x_{i-1}, x_i]$. If the function f actually was constant on these subintervals, the following Riemann sum would be the Riemann integral of $f(x)$ over $[a, b]$. In general however, each such sum is an approximation to the Riemann integral of $f(x)$ over $[a, b]$.

$$\sum_{i=1}^{n} f(x_i^\star)\Delta x_i. \tag{8.12}$$

Here $f(x_i^\star)$ is the sampled functional value from the subinterval $[x_{i-1}, x_i]$ and $\Delta x_i = x_i - x_{i-1}$ is the length of the i^{th} subinterval. As such the Riemann sum (8.12) is a finite sum of areas of rectangles.

A sequence of such Riemann sums is created by subdividing a previous partition and creating a new Riemann sum. The Riemann integral is, by definition, the limit of these Riemann sums as n, the number of subintervals tends to infinity with the lengths Δx_i tending to zero. One way to do this is to use a uniform partition, with $\Delta x_i = \frac{6-a}{n}$ for all n and all i. Elementary calculus texts often use this approach. It turns out that when such a sequence of Riemann sums has a limit, it is independent of the partitioning or the sampling. We define the Riemann integral of $f(x)$ over $[a, b]$ as

$$\int_a^b f(x)\,dx = \lim_{\Delta x_i \to 0} \sum_{i=1}^{n} f(x_i^\star)\Delta x_i. \tag{8.13}$$

We remark that this limit exists in particular if f is continuous on $[a, b]$, or bounded with a finite number of discontinuities. In summary, to define the Riemann integral, one partitions the domain of the functions by functional values. The Lebesgue integral reverses this point of view. In essence one partitions the range of the function and approximates the function

by a constant functional value over ***inverse images of the intervals partitioning the range.***

Definition 8. . ***Given a probability space*** $(\Omega, \mathcal{F}, P)$ ***and a function*** $F : \Omega \to \mathbb{R}$ ***(a random variable) we define the expectation, or average value of*** F***, as***

(8.) $$E(F) = \int_{\Omega} F(\omega) dP(\omega).$$

Notice that $P(\Omega) = 1$***. We also need conditional expectation of*** F***. This in general is another function.***

Definition 8. . ***Given a sub-*** σ***-algebra*** $\mathcal{G} \subset \mathcal{F}$***, we define*** $E(F|\,\mathcal{G})$***, the conditional expectation of*** F ***given*** $\mathcal{G}$***. This is any function*** G ***such that***

1. G ***is measurable with respect to the*** σ***-algebra*** $\mathcal{G}$***.***
2. $\int_A G dP = \int_A F dP$ ***for all*** $A \in \mathcal{G}$***.***

We remark that, in general, the smaller the σ***-algebra, the less functions are measurable. This is because less sets are available to allow the inverse images of measurable sets to be measurable.***

Definition 8. . ***Let*** $f(x) = x$ ***on*** $[0,1]$***. Now*** $([0,1], \mathcal{L}[0,1], m)$***, where*** $\mathcal{L}[0,1]$ ***are the Lebesgue measurable sets of*** $[0,1]$***, is a probability space. Notice that the function***

$$g = \begin{cases} \frac{1}{4}, & 0 \le x < \frac{1}{2} \\ \frac{3}{4}, & \frac{1}{2} \le x \le 1 \end{cases}$$

is measurable with respect to the σ***-algebra*** $\mathcal{G} = \sigma([0, \frac{1}{2}), [\frac{1}{2}, 1]) = \{\emptyset, (0, \frac{1}{2}), (\frac{1}{2}, 1], [0,1])$***, but that*** $f(x)$ ***is not. (For example,*** $\{x \in [0,1] | f(x) > \frac{3}{4}\}$ ***is not measurable with respect to*** $\mathcal{G}$***.)***

9. A Discussion of Brownian Motion

10. Derivation of the Black-Scholes Equation and Applications

References

Ikeda, N. and Watanabe, S., Stochastic Differential Equations and Diffusion Processes, North-Holland Publishing Company, Amsterdam, New York, Oxford, (October 28) 1981.

Durrett, R., Brownian Motion and Martingales in Analysis, Wadsworth, Inc., Belmont, California, 1984.

Hida, T., Brownian Motion, Springer-Verlaf, New York, (November 11) 1980.

Karatzas, I. and Shreve, S.E., Brownian Motion and Stochastic Calculus, Springer-Verlag, New York, (November 11) 1988.

Baxter, M. and Rennie, A., Financial Calculus,