MEASURE, INTEGRATION & PROBABILITY

Ivan F Wilde

Mathematics Department

King's College London

Chapter 1

σ-algebras and Borel functions

Preliminary discussion

• Suppose that X is a continuous random variable. Then $\text{Prob}(X = s) = 0$ for any $s \in \mathbb{R}$. However, the event $\{ X \in \mathbb{R} \}$ has probability one and is the disjoint union of the events $\{ X = s \}$ for $s \in \mathbb{R}$;

$$\{ X \in \mathbb{R} \} = \bigcup_s \{ X = s \}.$$

Each event on the right hand side has probability zero, so the probabilities of the events on the right hand side do not "add up" to that of the left hand side. We wish to understand this.

• Suppose that X is a random variable on a sample space Ω, and suppose that X takes values $x_1, x_2, \ldots$. Put $A_i = \{ \omega \in \Omega : X(\omega) = x_i \}$. Then

$$\begin{aligned} \mathbb{E}X &= \sum_i x_i \ \text{Prob}(X = x_i) \\ &= \sum_i x_i \ \text{Prob}(A_i). \end{aligned}$$

In particular, for $\mathbb{1}_{A_i}(\omega) = \begin{cases} 1, & \omega \in A_i \\ 0, & \omega \notin A_i \end{cases}$,

$$\begin{aligned} \mathbb{E}\,\mathbb{1}_{A_i} &= 1 \ \text{Prob}(\mathbb{1}_{A_i} = 1) \\ &= \text{Prob}(A_i). \end{aligned}$$

Also $\mathbb{E}(x_i \mathbb{1}_{A_i}) = x_i \ \text{Prob}(A_i)$. But we can write $X = \sum_i x_i \mathbb{1}_{A_i}$ and we recover $\mathbb{E}X$ as

$$\mathbb{E}X = \sum_i x_i \ \text{Prob}(A_i) = \sum_i \mathbb{E}(x_i \mathbb{1}_{A_i}).$$

Here X is a "step-function" on Ω. This formula forms the basis for the "general" expectation, i.e., that for an arbitrary random variable.

• One must (sometimes) ask which subsets of a sample space are deemed to be events. Can one take *all* subsets of the sample space to be events? The answer is sometimes yes and sometimes no. For example, in the case when the probability of an event within a bounded region of $\mathbb{R}^3$ is required to be proportional to the volume associated with the event, then one naturally asks whether every subset of (a bounded region) of $\mathbb{R}^3$ actually *has* a volume. That this is *not* so is demonstrated by the Banach-Tarski theorem. (This says that a ball of unit radius in $\mathbb{R}^3$ can be cut up into a finite number of pieces which can then be reassembled to form a ball of radius 2. The meaning of "volume" for these pieces is not clear.)

We must be precise about the concept of "event". In the "modern" (Kolmogorov) theory of probability, this is formulated in terms of σ-algebras.

Definition 1.1. A collection Σ of subsets of a non-empty set X is called a σ-algebra if

(i) $X \in \Sigma$,

(ii) if $A \in \Sigma$, then $A^c = X \setminus A \in \Sigma$,

(iii) if $A_n \in \Sigma$ for $n = 1, 2, \dots$, then $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

The sets in Σ are called measurable sets, and (X, Σ) is called a "measurable space".

Remarks 1.2.

1. Since $\varnothing = X^c$, it follows that $\varnothing \in \Sigma$.
2. For any $A_1, A_2, \dots, A_n \in \Sigma$, put $A_{n+1} = A_{n+2} = \cdots = \varnothing$. Then we see that $A_1 \cup \cdots \cup A_n = \bigcup_{k=1}^{\infty} A_k \in \Sigma$, by (1) above, and (iii).
3. Let $A_1, A_2, \dots \in \Sigma$. Then since $\bigcap_{n=1}^{\infty} A_n = \big(\bigcup_{n=1}^{\infty} A_n^c\big)^c$, we see that $\bigcap_{n=1}^{\infty} A_n \in \Sigma$.

   If we take $A_{n+1} = A_{n+2} = \cdots = X$, then we get $A_1 \cap \cdots \cap A_n \in \Sigma$.
4. Let $A, B \in \Sigma$. Then $A \setminus B = A \cap B^c$ and so $A \setminus B \in \Sigma$.

Proposition 1.3. *Let $\{\Sigma_\alpha\}$ be an arbitrary collection of σ-algebras of X. Then $\bigcap_\alpha \Sigma_\alpha$ is a σ-algebra.*

Proof. We check the requirements.

(i) $X \in \Sigma_\alpha$ for all α and so $X \in \bigcap_\alpha \Sigma_\alpha$.

(ii) Suppose that $A \in \bigcap_\alpha \Sigma_\alpha$. Then $A \in \Sigma_\alpha$ for all α and so $A^c \in \Sigma_\alpha$ for all α, that is $A^c \in \bigcap_\alpha \Sigma_\alpha$.

(iii) Let $A_1, A_2, \ldots$ belong to $\bigcap_\alpha \Sigma_\alpha$. Then, for each α, $A_n \in \Sigma_\alpha$ for all n and so $\bigcup_n A_n \in \Sigma_\alpha$. Hence $\bigcup_n A_n \in \bigcap_\alpha \Sigma_\alpha$.

The result follows. ∎

Let $\mathcal{C}$ be any collection of subsets of X. Then certainly $\mathcal{C}$ is contained in the σ-algebra consisting of *all* subsets of X. If we set

$$\Sigma(\mathcal{C}) = \bigcap_{\mathcal{F}} \Sigma$$

where the intersection is over the family $\mathcal{F}$ of all those σ-algebras Σ which contain $\mathcal{C}$, then $\Sigma(\mathcal{C})$ is the "smallest" σ-algebra containing $\mathcal{C}$. It is called the σ-algebra generated by $\mathcal{C}$.

Definition 1.4. Let $\mathcal{C}$ denote the collection of open subsets of $\mathbb{R}$. Then $\Sigma(\mathcal{C})$ is called the Borel σ-algebra of $\mathbb{R}$, usually written $\mathcal{B}(\mathbb{R})$. The elements of $\mathcal{B}(\mathbb{R})$ are called Borel sets. Similarly, one defines $\mathcal{B}(\mathbb{R}^n)$ as the σ-algebra generated by the open subsets of $\mathbb{R}^n$.

Proposition 1.5. *The following subsets of* $\mathbb{R}$ *belong to* $\mathcal{B}(\mathbb{R})$*:*

(i) (a,b) *for any* $a < b$;

(ii) $(-\infty, a)$ *for any* $a \in \mathbb{R}$;

(iii) (a, ∞) *for any* $a \in \mathbb{R}$;

(iv) $[a,b]$ *for any* $a \le b$;

(v) $(-\infty, a]$ *for any* $a \in \mathbb{R}$;

(vi) $[a, \infty)$ *for any* $a \in \mathbb{R}$;

(vii) $(a,b]$ *for any* $a < b$;

(viii) $[a,b)$ *for any* $a < b$;

(ix) *any closed subset of* $\mathbb{R}$.

Proof. Each of the sets in (i), (ii) or (iii) is open and so belongs to $\mathcal{B}(\mathbb{R})$ by construction.

(iv) $[a,b] = \bigcap_{n=1}^\infty \left(a - \frac{1}{n}, b + \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R})$.

(v) $(-\infty, a] = \bigcap_{n=1}^\infty \left(-\infty, a + \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R})$.

(vi) $[a, \infty) = \bigcap_{n=1}^\infty \left(a - \frac{1}{n}, \infty\right) \in \mathcal{B}(\mathbb{R})$.

(vii) $(a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R})$.

(viii) $[a, b) = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right) \in \mathcal{B}(\mathbb{R})$.

(ix) If F is closed in $\mathbb{R}$, then F^c is open and so belongs to $\mathcal{B}(\mathbb{R})$. But then $F = (F^c)^c \in \mathcal{B}(\mathbb{R})$. ∎

In fact, we will se that each of these families of subsets of $\mathbb{R}$ generates the σ-algebra $\mathcal{B}(\mathbb{R})$.

Proposition 1.6. *Let Σ(closed) and Σ(compact) denote the σ-algebras of subsets of $\mathbb{R}$ generated, respectively, by the closed sets and the compact subsets in $\mathbb{R}$. Then*

$$\Sigma(\textit{closed}) = \Sigma(\textit{compact}) = \mathcal{B}(\mathbb{R}).$$

Proof. Every closed subset of $\mathbb{R}$ belongs to the σ-algebra $\mathcal{B}(\mathbb{R})$. But, by definition, Σ(closed) is the smallest σ-algebra containing the closed sets, so we must have $\Sigma(\text{closed}) \subseteq \mathcal{B}(\mathbb{R})$. On the other hand, every open set is the complement of a closed set and so belongs to Σ(closed). By definition, $\mathcal{B}(\mathbb{R})$ is the σ-algebra generated by the open sets of $\mathbb{R}$ and so we have $\mathcal{B}(\mathbb{R}) \subseteq \Sigma(\text{closed})$. It follows that $\mathcal{B}(\mathbb{R}) = \Sigma(\text{closed})$.

Next, we note that since every compact set in $\mathbb{R}$ is closed, it follows that $\Sigma(\text{compact}) \subseteq \Sigma(\text{closed})$. However, any closed set F can be written as the (countable) union

$$F = \bigcup_{n=1}^{\infty} [-n, n] \cap F.$$

Each $[-n, n] \cap F$ is closed and bounded and so is compact. It follows that $F \in \Sigma(\text{compact})$ and therefore $\Sigma(\text{closed}) \subseteq \Sigma(\text{compact})$. Hence result. ∎

Proposition 1.7. *Let $\mathcal{C}_1, \dots, \mathcal{C}_9$ denote the collections of subsets of $\mathbb{R}$ as given in Proposition 1.5. Then $\Sigma(\mathcal{C}_i) = \mathcal{B}(\mathbb{R})$ for each $i = 1, 2, \dots, 9$.*

Proof. Since $\mathcal{C}_i \subseteq \mathcal{B}(\mathbb{R})$, we have $\Sigma(\mathcal{C}_i) \subseteq \mathcal{B}(\mathbb{R})$, $1 \le i \le 9$. We show that $\Sigma(\text{compact}) \subseteq \Sigma(\mathcal{C}_i)$ which completes the proof, by Proposition 1.6.

To show this, we first observe that each $\Sigma(\mathcal{C}_i)$ contains all intervals (a, b) with $a < b$. (For example in (v): $(-\infty, a] \in \Sigma(\mathcal{C}_5)$ implies (by taking complements) that $(a, \infty) \in \Sigma(\mathcal{C}_5)$ for any $a \in \mathbb{R}$. But then it follows that $(a, b] = (a, \infty) \cap (-\infty, b] \in \Sigma(\mathcal{C}_5)$ and so $(a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{b-a}{2n}\right] \in \Sigma(\mathcal{C}_5)$.)

Now let $K \subset \mathbb{R}$ be any given compact set. For each $x \in \mathbb{R}$ and $n \in \mathbb{N}$, let $I_n(x)$ be the interval $(x - \frac{1}{n}, x + \frac{1}{n})$. For each fixed n, the collection $\{ I_n(x) : x \in K \}$ is an open cover of K and so has a finite subcover,

$$K \subset I_n(x_1^{(n)}) \cup \cdots \cup I_n(x_{m(n)}^{(n)}) \equiv J_n$$

for suitable points $x_1^{(n)}, \dots, x_{m(n)}^{(n)}$ in K. Evidently, $J_n \in \Sigma(\mathcal{C}_i)$.

We claim that $K = \bigcap_{n=1}^{\infty} J_n$. Clearly, $K \subseteq \bigcap_{n=1}^{\infty} J_n$. For the converse, let $x \in \bigcap_{n=1}^{\infty} J_n$. Then $x \in J_n$ for all n. Hence x belongs to some $I_n(x_j^{(n)})$, that is, there is $y_n = x_j^{(n)} \in K$ such that $|x - y_n| < 1/n$. We see that $y_n \to x$ as $n \to \infty$ and, since K is closed, it follows that $x \in K$ which proves the claim. Therefore $K = \bigcap_{n=1}^{\infty} J_n \in \Sigma(\mathcal{C}_i)$ and so

$$\mathcal{B}(\mathbb{R}) = \Sigma(\text{compact}) \subseteq \Sigma(\mathcal{C}_i)$$

and the proof is complete. ∎

Definition 1.8. Let (X, Σ) be a measurable space and $f : X \to \mathbb{R}$ a given function. f is said to be Borel measurable if $f^{-1}(G) \in \Sigma$ for each G open in $\mathbb{R}$.

Proposition 1.9. *The function $f : X \to \mathbb{R}$ is Borel measurable if and only if $f^{-1}(A) \in \Sigma$ for each $A \in \mathcal{B}(\mathbb{R})$.*

Proof. If $f^{-1}(A) \in \Sigma$ for each $A \in \mathcal{B}(\mathbb{R})$, then certainly $f^{-1}(G) \in \Sigma$ for each open set G in $\mathbb{R}$ (because such G belongs to $\mathcal{B}(\mathbb{R})$).

Conversely, suppose that $f^{-1}(G) \in \Sigma$ for any open set G in $\mathbb{R}$. Let $\mathcal{S}$ denote the collection of subsets of $\mathbb{R}$ given by

$$\mathcal{S} = \{ E \subseteq \mathbb{R} : f^{-1}(E) \in \Sigma \}.$$

Then $\mathcal{S}$ is a σ-algebra. To see this, we note the following.

(i) $f^{-1}(\mathbb{R}) = X \in \Sigma$, so $\mathbb{R} \in \mathcal{S}$.

(ii) $f^{-1}(\mathbb{R} \setminus E) = X \setminus f^{-1}(E)$, so if $E \in \mathcal{S}$ then so is $\mathbb{R} \setminus E$.

(iii) If $E_1, E_2, \dots$ belong to $\mathcal{S}$, then

$$f^{-1}(E_1 \cup E_2 \cup \dots) = f^{-1}(E_1) \cup f^{-1}(E_2) \cup \dots$$

which belongs to Σ and so $E_1 \cup E_2 \cup \cdots \in \mathcal{S}$.

This shows that $\mathcal{S}$ is, indeed, a σ-algebra, as claimed. But $\mathcal{S}$ contains all open sets, by hypothesis, and therefore $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{S}$. Hence, for any $A \in \mathcal{B}(\mathbb{R})$, $f^{-1}(A) \in \Sigma$. This completes the proof. ∎

Remark 1.10. Note that $\mathcal{S}$ need not be equal to $\mathcal{B}(\mathbb{R})$. For example, if X is the σ-algebra of all subsets of X then every function $f : X \to \mathbb{R}$ is Borel measurable and $f^{-1}(E) \in \Sigma$ for any subset E in $\mathbb{R}$ whatsoever.

As another example, suppose that f is constant. Then $f^{-1}(E)$ is either equal to X or else is empty depending on whether E contains the value assumed by f or not. In any event, $f^{-1}(E) \in \Sigma$, whatever E is.

We can improve somewhat on the previous proposition, still using the same idea.

Proposition 1.11. *Let $\mathcal{C}$ be a collection of subsets of $\mathbb{R}$ such that $\Sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$ and let $f : X \to \mathbb{R}$. Then f is Borel measurable if and only if $f^{-1}(A) \in \Sigma$ for all $A \in \mathcal{C}$.*

Proof. Suppose that f is Borel measurable. Since $\mathcal{C} \subseteq \mathcal{B}(\mathbb{R})$, it follows by Proposition 1.9 that $f^{-1}(A) \in \Sigma$ for any $A \in \mathcal{C}$.

Now suppose that $f^{-1}(A) \in \Sigma$ for all $A \in \mathcal{C}$. As before, let us set $\mathcal{S} = \{ E \subseteq \mathbb{R} : f^{-1}(E) \in \Sigma \}$. Then $\mathcal{S}$ is a σ-algebra which contains $\mathcal{C}$ and so we have

$$\mathcal{B}(\mathbb{R}) = \Sigma(\mathcal{C}) \subseteq \mathcal{S}.$$

It follows that $f^{-1}(A) \in \Sigma$ for any $A \in \mathcal{B}(\mathbb{R})$. ■

Remark 1.12. This means that we can take $\mathcal{C}$ to be any of the collections of sets indicated in Proposition 1.5. For example, we can say that f is Borel measurable if and only if $f^{-1}((-\infty, a]) \in \Sigma$ for each $a \in \mathbb{R}$.

We can choose any convenient collection to work with.

Proposition 1.13. *Let $f : X \to \mathbb{R}$ be Borel measurable and $g : \mathbb{R} \to \mathbb{R}$ continuous. Then $g \circ f : X \to \mathbb{R}$ is Borel measurable.*

Proof. Let G be any open set in $\mathbb{R}$. Then $g^{-1}(G)$ is open in $\mathbb{R}$ and so $(g \circ f)^{-1}(G) = f^{-1}(g^{-1}(G))$ belongs to Σ. ■

Proposition 1.14. *Let $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ be Borel functions. Then the set*

$$E = \{ x \in X : f(x) < g(x) \}$$

is measurable.

Proof. For each rational number $r \in \mathbb{Q}$, let

$$E_r = \{ x \in X : f(x) < r < g(x) \}.$$

Then $E_r = \{ x : f(x) < r \} \cap \{ x : r < g(x) \}$ which is the intersection of two measurable sets in X and so is itself measurable. Finally, we note that $E = \bigcup_{r \in \mathbb{Q}} E_r$ which is a countable union of measurable sets and so is measurable. ■

Proposition 1.15. *Let (X, Σ) be a measurable space and let $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ be Borel functions. Then*

(i) *$af + b$ is a Borel function for any $a, b \in \mathbb{R}$;*

(ii) *$f + g$ is a Borel function;*

(iii) $|f|^\alpha$ *is a Borel function for any* $\alpha \ge 0$*;*

(iv) *if* f *never vanishes, then* $1/f$ *is a Borel function;*

(v) fg *is a Borel function;*

(vi) $|f|$, $\max\{\, f, g \,\}$ *and* $\min\{\, f, g \,\}$ *are Borel functions.*

Proof. We shall consider each statement one by one.
(i) For any $c \in \mathbb{R}$,

$$\begin{aligned}
\{\, x \in X : (af + b)(x) \le c \,\} &= \{\, x : a\,f(x) + b \le c \,\} \\
&= \{\, x : a\,f(x) \le c - b \,\} \\
&= \begin{cases} \{\, x : f(x) \le (c-b)/a \,\}, & a > 0 \\ \{\, x : f(x) \ge (c-b)/a \,\}, & a < 0 \\ \quad X, & a = 0 \text{ and } c \ge b \\ \quad \varnothing, & a = 0 \text{ and } c < b. \end{cases}
\end{aligned}$$

In any event, the left hand side belongs to Σ, which proves (i).

(ii) Let $c \in \mathbb{R}$. Then

$$\{\, x : f(x) + g(x) > c \,\} = \{\, x : f(x) > -g(x) + c \,\}.$$

But $-g + c$ is a Borel function by (i) and so the right hand side belongs to Σ by Proposition 1.14.

(iii) The function $t \mapsto |t|^\alpha$ is a continuous function and so $|f|^\alpha$ is a Borel function by Proposition 1.13. (Alternatively, we note that for $c \ge 0$

$$\begin{aligned}
\{\, x : |f(x)|^\alpha \le c \,\} &= \{\, x : -c^{1/\alpha} \le f(x) \le c^{1/\alpha} \,\} \\
&= \{\, x : -c^{1/\alpha} \le f(x) \,\} \cap \{\, x : f(x) \le c^{1/\alpha} \,\} \\
&\in \Sigma.
\end{aligned}$$

For $c < 0$, the left hand side $= \varnothing$.

(iv) If $c \ge 0$, then

$$\{\, x : 1/f(x) \le c \,\} = \{\, x : f(x) < 0 \,\} \cup \{\, x : 1 \ge c\,f(x) \,\} \in \Sigma,$$

using (i). If $c < 0$, then

$$\{\, x : 1/f(x) \le c \,\} = \{\, x : 1 \ge c\,f(x) \,\} \in \Sigma,$$

again by (i).

(v) This follows from the identity

$$fg = \tfrac{1}{4}(f+g)^2 - \tfrac{1}{4}(f-g)^2$$

together with (i), (ii) and (iii).

(vi) $|f|$ is a Borel function by (iii) with $\alpha = 1$. Now

$$\max\{\, f, g \,\}(x) = \tfrac{1}{2}(f(x)+g(x)) + \tfrac{1}{2}\,|f(x)-g(x)|$$

and

$$\min\{\, f, g \,\}(x) = \tfrac{1}{2}(f(x)+g(x)) - \tfrac{1}{2}\,|f(x)-g(x)|\,.$$

The result now follows from (i), (ii) and (iii). ∎

Theorem 1.16. *Let (X,Σ) be a measurable space and let (f_n) be a sequence of Borel measurable functions on X. Suppose that $f(x) = \lim_n f_n(x)$ exists for each $x \in X$. Then f is a Borel function.*

Proof. Let $c \in \mathbb{R}$. We shall show that $A = \{\, x : f(x) < c \,\} \in \Sigma$. For any $m, k \in \mathbb{N}$, put

$$E_k^m = \{\, x : f_n(x) < c - \tfrac{1}{m} \text{ for all } n > k \,\}.$$

Then $E_k^m = \bigcap_{n>k} \{\, x : f_n(x) < c - \frac{1}{m} \,\}$ and so $E_k^m \in \Sigma$.

Claim: $A = \displaystyle\bigcup_{m=1}^{\infty} \bigcup_{k=1}^{\infty} E_k^m$.

To see this, fix $x \in A$. Then $f(x) < c$, by definition of A and there is some $m_0 \in \mathbb{N}$ such that $f(x) < c - \frac{1}{m_0}$. But $f_n(x) \to f(x)$ and so there is some $k_0 \in \mathbb{N}$ such that $f_n(x) < c - \frac{1}{m_0}$ for all $n > k_0$. In other words, $x \in E_{k_0}^{m_0}$.

Now suppose that $x \in E_k^m$ for some m, k. Then $f_n(x) < c - \frac{1}{m}$ for all $n > k$. In particular, $f(x) = \lim_n f_n(x) \le c - \frac{1}{m} < c$ and so $x \in A$. This proves the claim.

Each $E_k^m \in \Sigma$ and so $A = \displaystyle\bigcup_{m=1}^{\infty} \bigcup_{k=1}^{\infty} E_k^m = \bigcup_{(m,k)\in\mathbb{N}^2} E_k^m$ is a countable union of elements of Σ and so itself belongs to Σ. ∎

Definition 1.17. Let (X,Σ) be a measurable space and $f : X \to \mathbb{C}$. We say that f is Borel measurable if both $\operatorname{Re} f$ and $\operatorname{Im} f$ are Borel measurable.

Proposition 1.18. *Let $f : X \to \mathbb{C}$ be a Borel measurable function. Then $|f|$ is a Borel measurable function and there is a Borel function $\alpha : X \to \mathbb{C}$ with $|\alpha(x)| = 1$ for all $x \in X$ such that $f(x) = \alpha(x)\,|f(x)|$.*

Proof. For each x, $f(x) = |f(x)|\, e^{i\theta}$. The value of θ depends on x and is arbitrary up to constant additions of multiples of 2π, which may vary with x. The point is that θ may be chosen such that the resulting function $e^{i\theta(x)}$ is Borel measurable.

First, we note that $|f(x)| = \big((\operatorname{Re} f(x))^2 + (\operatorname{Im} f(x))^2\big)^{1/2}$ and so is Borel measurable, by Proposition 1.15.

Let $E = \{\, x : f(x) = 0 \,\} = \{\, x : |f(x)| = 0 \,\}$. Then $E \in \Sigma$. Let 1_E be the indicator function for E,

$$1_E(x) = \begin{cases} 1, & x \in E, \\ 0, & \text{otherwise.} \end{cases}$$

Then 1_E is a Borel function.

Set

$$\alpha(x) = \frac{f(x) + 1_E(x)}{|f(x)| + 1_E(x)}\,, \quad \text{for } x \in X.$$

Then, evidently, $|\alpha(x)| = 1$ (consider separately the cases $x \in E$ and $x \notin E$). Furthermore, $|f| + 1_E$ never vanishes and is a Borel function, and so therefore is $1/(|f| + 1_E)$, by Proposition 1.15. Hence we deduce, again by Proposition 1.15, that α is a complex Borel function. It is clear that $f(x) = \alpha(x)\,|f(x)|$ for all $x \in X$. ■

Definition 1.19. A function $s : X \to \mathbb{R}$ is called a simple function if its range consists of only finitely-many points.

Thus, if s is simple with distinct values $\alpha_1, \dots, \alpha_n$ and if $A_j = \{\, x \in X : s(x) = \alpha_j \,\}$, $1 \le j \le n$, then

$$s(x) = \sum_{j=1}^{n} \alpha_j\, 1_{A_j}(x)$$

for $x \in X$, i.e., s is a finite linear combination of indicator functions. Evidently, s is Borel if and only if each A_j is measurable.

Theorem 1.20. *Let $f : X \to \mathbb{R}$ be a non-negative Borel function. Then there is a sequence of non-negative simple Borel functions (s_n) such that*

(i) $0 \le s_1 \le s_2 \le \cdots \le f$,

(ii) $s_n(x) \to f(x)$, *as $n \to \infty$, for each $x \in X$.*

Proof. For $n = 1, 2, \dots$ and for $1 \le i \le n2^n$, let

$$E_{n,i} = \{\, x \in X : \frac{i-1}{2^n} \le f(x) < \frac{i}{2^n} \,\}$$

and

$$F_n = \{\, x \in X : f(x) \ge n \,\}.$$

Then $F_n \in \Sigma$ and $E_{n,i} \in \Sigma$ and $X = F_n \cup \bigcup_{i=1}^{n2^n} E_{n,i}$.

Set

$$s_n(x) = \sum_{i=1}^{n2^n} \frac{(i-1)}{2^n} \, 1_{E_{n,i}}(x) + n 1_{F_n}(x).$$

Then s_n is a non-negative simple (Borel) function obeying $s_n(x) \le f(x)$ for each $x \in X$.

To show that $s_n \le s_{n+1}$, suppose first that $x \in F_{n+1}$. Then we see that $s_{n+1}(x) = n + 1 > n = s_n(x)$ since $F_{n+1} \subseteq F_n$.

Now, if $x \notin F_{n+1}$, there is some j with $1 \le j \le (n+1)2^{n+1}$ such that $x \in E_{n+1,j}$. Then

$$s_{n+1}(x) = \frac{j-1}{2^{n+1}} \equiv \frac{[2^{n+1} f(x)]}{2^{n+1}}$$

where $[v]$ denotes the integer part of v. (This last equality follows because $x \in E_{n+1,j}$ if and only if $(j-1)/2^{n+1} \le f(x) < j/2^{n+1}$ which holds if and only if $j - 1 \le 2^{n+1} f(x) < j$ which, in turn, holds if and only if $j - 1 = [2^{n+1} f(x)]$.)

If $f(x) \ge n$, then $[2^{n+1} f(x)] \ge 2^{n+1} n$ and so

$$s_{n+1}(x) = \frac{[2^{n+1} f(x)]}{2^{n+1}} \ge \frac{2^{n+1} n}{2^{n+1}} = n = s_n(x).$$

Suppose now that $0 \le f(x) < n$. Then $x \in E_{n,i}$ for some $0 \le i \le n2^n$ and $s_n(x) = [2^n f(x)]/2^n$.

Suppose that $2^n f(x) = m + \lambda$, some $m = 0, 1, 2, \dots$ and $0 \le \lambda < 1$, so that $[2^n f(x)] = m$. Then $2^{n+1} f(x) = 2m + 2\lambda$ with $0 \le 2\lambda < 2$ giving

$$\begin{aligned} \frac{[2^{n+1} f(x)]}{2^{n+1}} &= \frac{2m + [2\lambda]}{2^{n+1}} \\ &= \frac{m}{2^n} + \frac{[2\lambda]}{2^{n+1}} \\ &\ge \frac{m}{2^n} \\ &= \frac{[2^n f(x)]}{2^n} \\ &= s_n(x), \end{aligned}$$

which completes the proof of (i).

To see that $s_n(x) \to f(x)$ for each x, let $x \in X$ be given and let n_0 be so large that $f(x) < n_0$, i.e., $x \notin F_n$ for any $n \ge n_0$. Hence, for all $n \ge n_0$, there is some i (depending on n) such that $x \in E_{n,i}$. Therefore, by definition of $E_{n,i}$,

$$0 \le f(x) - s_n(x) < \frac{i}{2^n} - \frac{(i-1)}{2^n}$$

$$= \frac{1}{2^n}$$
$$\to 0, \quad \text{as } n \to \infty,$$

and the proof of (ii) is complete. ■

Remark 1.21. If f is bounded, then there is n_0 such that $0 \le f(x) < n_0$ for all $x \in X$. But then for all $n > n_0$, $0 \le f(x) - s_n(x) < 1/2^n$, for any $x \in X$, and so we see that $s_n \to f$ uniformly on X in this case.

Definition 1.22. A collection $\mathcal{A}$ of subsets of a set X is an algebra if

(i) $X \in \mathcal{A}$,

(ii) if $A \in \mathcal{A}$ and $B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$,

(iii) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.

Note that it follows that if $A, B \in \mathcal{A}$, then $A \cap B = (A^c \cup B^c)^c \in \mathcal{A}$. Also, for any finite family $A_1, \dots, A_n \in \mathcal{A}$, it follows by induction that $\bigcup_{i=1}^n A_i \in \mathcal{A}$ and $\bigcap_{i=1}^n A_i \in \mathcal{A}$.

Definition 1.23. A collection $\mathcal{M}$ of subsets of X is a monotone class if

(i) whenever $A_1 \subseteq A_2 \subseteq \dots$ in $\mathcal{M}$ is an increasing sequence in $\mathcal{M}$, then $\bigcup_{i=1}^\infty A_i \in \mathcal{M}$,

(ii) whenever $B_1 \supseteq B_2 \supseteq \dots$ in $\mathcal{M}$ is a decreasing sequence in $\mathcal{M}$, then $\bigcap_{i=1}^\infty B_i \in \mathcal{M}$.

One can show that the intersection of an arbitrary family of monotone classes of subsets of X is itself a monotone class. Thus, given collection $\mathcal{C}$ of subsets of X, we may consider $\mathcal{M}(\mathcal{C})$, the monotone class generated by the collection $\mathcal{C}$ — it is the "smallest" monotone class containing $\mathcal{C}$, i.e., it is the intersection of all those monotone classes which contain $\mathcal{C}$.

Theorem 1.24. *Let $\mathcal{A}$ be an algebra of subsets of X. Then $\mathcal{M}(\mathcal{A}) = \Sigma(\mathcal{A})$.*

Proof. It is clear that any σ-algebra is a monotone class and so $\Sigma(\mathcal{A})$ is a monotone class containing $\mathcal{A}$. Hence $\mathcal{M}(\mathcal{A}) \subseteq \Sigma(\mathcal{A})$. The proof is complete if we can show that $\mathcal{M}(\mathcal{A})$ is a σ-algebra, for then we would deduce that $\Sigma(\mathcal{A}) \subseteq \mathcal{M}(\mathcal{A})$.

If a monotone class $\mathcal{M}$ is an algebra, then it is a σ-algebra. To see this, let $A_1, A_2, \dots \in \mathcal{M}$. For each $n \in \mathbb{N}$, set $B_n = A_1 \cup \dots \cup A_n$. Then $B_n \in \mathcal{M}$, if $\mathcal{M}$ is an algebra. But then $\bigcup_{i=1}^\infty A_i = \bigcup_{n=1} \infty B_n \in \mathcal{M}$ if $\mathcal{M}$ is a monotone class. Thus the algebra $\mathcal{M}$ is also a σ-algebra. It remains to prove that $\mathcal{M}$ is, in fact, an algebra. We shall verify the three requirements.
(i) We have $X \in \mathcal{A} \subseteq \mathcal{M}(\mathcal{A})$.

(iii) Let $A \in \mathcal{M}(\mathcal{A})$. We wish to show that $A^c \in \mathcal{M}(\mathcal{A})$. To show this, let

$$\widetilde{\mathcal{M}} = \{ B : B \in \mathcal{M}(\mathcal{A}) \text{ and } B^c \in \mathcal{M}(\mathcal{A}) \}.$$

Since $\mathcal{A}$ is an algebra, if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$ and so

$$\mathcal{A} \subseteq \widetilde{\mathcal{M}} \subseteq \mathcal{M}(\mathcal{A}).$$

We shall show that $\widetilde{\mathcal{M}}$ is a monotone class. Let (B_n) be a sequence in $\widetilde{\mathcal{M}}$ with $B_1 \subseteq B_2 \subseteq \dots$. Then $B_n \in \mathcal{M}(\mathcal{A})$ and $B_n^c \in \mathcal{M}(\mathcal{A})$. Hence $\bigcup_n B_n \in \mathcal{M}(\mathcal{A})$ and also $\bigcap_n B_n^c \in \mathcal{M}(\mathcal{A})$, since $\mathcal{M}(\mathcal{A})$ is a monotone class (and (B_n^c) is a decreasing sequence).

But $\bigcap_n B_n^c = \big(\bigcup_n B_n\big)^c$ and so both $\bigcup_n B_n$ and $\big(\bigcup_n B_n\big)^c$ belong to $\mathcal{M}(\mathcal{A})$, i.e., $\bigcup_n B_n \in \widetilde{\mathcal{M}}$.

Similarly, if $B_1 \supseteq B_2 \supseteq \dots$ belong to $\widetilde{\mathcal{M}}$, then $\bigcap_n B_n \in \mathcal{M}(\mathcal{A})$ and $\big(\bigcap_n B_n\big)^c = \bigcup_n B_n^c \in \mathcal{M}(\mathcal{A})$ so that $\bigcap_n B_n \in \widetilde{\mathcal{M}}$. It follows that $\widetilde{\mathcal{M}}$ is a monotone class. Since $\mathcal{A} \subseteq \widetilde{\mathcal{M}} \subseteq \mathcal{M}(\mathcal{A})$ and $\mathcal{M}(\mathcal{A})$ is the monotone class generated by $\mathcal{A}$, we conclude that $\widetilde{\mathcal{M}} = \mathcal{M}(\mathcal{A})$. But then this means that for any $B \in \mathcal{M}(\mathcal{A})$, we also have $B^c \in \mathcal{M}(\mathcal{A})$.

(ii) We wish to show that if A and B belong to $\mathcal{M}(\mathcal{A})$ then so does $A \cup B$. By (iii), it is enough to show that $A \cap B \in \mathcal{M}(\mathcal{A})$ (using $A \cup B = (A^c \cap B^c)^c$). To this end, let $A \in \mathcal{M}(\mathcal{A})$ and let

$$\mathcal{M}_A = \{ B : B \in \mathcal{M}(\mathcal{A}) \text{ and } A \cap B \in \mathcal{M}(\mathcal{A}) \}.$$

Then for $B_1 \subseteq B_2 \subseteq \dots$ in $\mathcal{M}_A$, we have

$$A \cap \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A \cap B_i \in \mathcal{M}(\mathcal{A})$$

since each $A \cap B_i \in \mathcal{M}(\mathcal{A})$ by the definition of $\mathcal{M}_A$.

Similarly, if $B_1 \supseteq B_2 \supseteq \dots$ belong to $\mathcal{M}_A$, then

$$A \cap \bigcap_{i=1}^{\infty} B_i = \bigcap_{i=1}^{\infty} A \cap B_i \in \mathcal{M}(\mathcal{A}).$$

Therefore $\mathcal{M}_A$ is a monotone class.

Suppose $A \in \mathcal{A}$. Then for any $B \in \mathcal{A}$, we have $A \cap B \in \mathcal{A}$, since $\mathcal{A}$ is an algebra. Hence $\mathcal{A} \subseteq \mathcal{M}_A \subseteq \mathcal{M}(\mathcal{A})$ and therefore $\mathcal{M}_A = \mathcal{M}(\mathcal{A})$ for each $A \in \mathcal{A}$.

Now, for any $B \in \mathcal{M}(\mathcal{A})$ and $A \in \mathcal{A}$, we have

$$A \in \mathcal{M}_B \iff A \cap B \in \mathcal{M}(\mathcal{A}) \iff B \in \mathcal{M}_A = \mathcal{M}(\mathcal{A}).$$

Hence, for every $B \in \mathcal{M}(\mathcal{A})$,

$$\mathcal{A} \subseteq \mathcal{M}_B \subseteq \mathcal{M}(\mathcal{A})$$

and so (since $\mathcal{M}_B$ is a monotone class) we have $\mathcal{M}_B = \mathcal{M}(\mathcal{A})$ for every $B \in \mathcal{M}(\mathcal{A})$.

Now let $A, B \in \mathcal{M}(\mathcal{A})$. We have seen that $\mathcal{M}_B = \mathcal{M}(\mathcal{A})$ and therefore $A \in \mathcal{M}(\mathcal{A})$ means that $A \in \mathcal{M}_B$ so that $A \cap B \in \mathcal{M}(\mathcal{A})$ and the proof is complete. ■

Chapter 2

Measures

Definition 2.1. A finite measure on a measurable space (X, Σ) is a map $\mu : \Sigma \to [0, \infty)$ such that if $A_1, A_2, \ldots$ is any sequence of pairwise disjoint members of Σ, then

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

(This requirement is referred to as "countable-additivity" or "σ-additivity".)

A measure space is a triple (X, Σ, μ), where μ is a measure on the σ-algebra Σ of subsets of X.

If $\mu(X) = 1$, then μ is called a probability measure and (X, Σ, μ) is called a probability space. In this case, X is called the sample space and the members of Σ are called events.

A random variable is a Borel measurable function on a probability space.

Proposition 2.2. *Let μ be a finite measure on Σ. Then the following hold.*

(i) $\mu(\varnothing) = 0$.

(ii) *If $A_1, \ldots, A_n \in \Sigma$ with $A_i \cap A_j = \varnothing$ for $i \neq j$, then*

$$\mu(A_1 + A_2 + \cdots + A_n) = \mu(A_1) + \cdots + \mu(A_n).$$

(iii) *If $A, B \in \Sigma$ with $A \subseteq B$, then $\mu(A) \le \mu(B)$.*

(iv) *If $A_1 \subseteq A_2 \subseteq \ldots$ with $A_n \in \Sigma$, $n = 1, 2, \ldots$, then we have $\mu(A_n) \uparrow \mu(\bigcup_m A_m)$ as $n \to \infty$.*

(v) *If $A_1 \supseteq A_2 \supseteq \ldots$ with $A_n \in \Sigma$ for $n = 1, 2, \ldots$, then we have $\mu(A_n) \downarrow \mu(\bigcap_m A_m)$ as $n \to \infty$.*

Proof. (i) Let $A_n = \varnothing$ for each $n \in \mathbb{N}$. Then the A_ns are pairwise disjoint and so $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$, that is, $\mu(\varnothing) = \sum_n \mu(\varnothing)$. We must have $\mu(\varnothing) = 0$.

(ii) Set $A_k = \varnothing$ for all $k > n$ and use (i) together with the countable additivity of μ.

(iii) We have $B = A \cup (B \setminus A)$ and $A \cap (B \setminus A) = \varnothing$. Hence, using (ii),

$$\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A).$$

(iv) Put $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2$, $\dots$. Then $B_n \in \Sigma$ and $B_i \cap B_j = \varnothing$ for all $i \neq j$. Also $A_n = B_1 \cup \dots \cup B_n$ and so $\bigcup_m A_m = \bigcup B_m$. Hence

$$\mu(A_n) = \sum_{i=1}^{n} \mu(B_i), \quad \text{by (ii)},$$

and $\mu(\bigcup_m A_m) = \mu(\bigcup_m B_m)$. But

$$\begin{aligned}
\mu(\bigcup_m B_m) &= \sum_{m=1}^{\infty} \mu(B_m), \quad \text{by } \sigma\text{-additivity}, \\
&= \lim_n \sum_{m=1}^{n} \mu(B_m) \\
&= \lim_n \mu(A_n).
\end{aligned}$$

Since $A_n \subseteq A_{n+1}$, $(\mu(A_n))$ is an increasing sequence.

(v) Set $C_n = A_1 \setminus A_n$. Then $C_n \in \Sigma$ and $C_1 \subseteq C_2 \subseteq \dots$. Let $A = \bigcap_m A_m$. Then $A_1 \setminus A = \bigcup_n C_n$ giving

$$\mu(A_1 \setminus A) = \lim_n \mu(C_n), \quad \text{by (iv)},$$

But $\mu(C_n) = \mu(A_1 \setminus A_n) = \mu(A_1) - \mu(A_n)$ and so

$$\mu(A_1) - \mu(A) = \lim_n (\mu(A_1) - \mu(A_n))$$

giving $\mu(A) = \lim_n \mu(A_n)$. The sequence $(\mu(A_n))$ is decreasing since we have $A_{n+1} \subseteq A_n$. ■

Proposition 2.3. *Suppose that $\mu : \Sigma \to [0, \infty)$ and that $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \in \Sigma$ with $A \cap B = \varnothing$ (i.e., μ is finitely-additive). Then μ is σ-additive if and only if $\mu(E_n) \downarrow 0$ for every sequence (E_n) in Σ satisfying $E_1 \supseteq E_2 \supseteq \dots$ and $\bigcap_n E_n = \varnothing$.*

Proof. Suppose that μ is σ-additive. Then by Proposition 2.2 (part (v)), we see that if E_n is a decreasing sequence in Σ with $\bigcap_n E_n = \varnothing$, then $\mu(E_n) \downarrow \mu(\varnothing) = 0$.

Conversely, suppose $\mu(E_n) \downarrow 0$ for any sequence (E_n) in Σ as above. Let A_n be any given sequence in Σ such that $A_i \cap A_j = \varnothing$ for any $i \neq j$. Let

$A = \bigcup_n A_n$ and set $E_n = A \setminus (A_1 \cup \dots \cup A_n)$ for $n \in \mathbb{N}$. Evidently, $E_n \in \Sigma$, $E_n \supseteq E_{n+1}$ and $\bigcap_n E_n = \varnothing$ and so, by hypothesis, $\mu(E_n) \downarrow 0$. However,

$$\mu(E_n) = \mu(A) - \mu(A_1 \cup \dots \cup A_n) = \mu(A) - \sum_{i=1}^{n} \mu(A_i)$$

since μ is finitely additive. We conclude that $\lim_n \sum_{i=1}^n \mu(A_i) = \mu(A)$, that is, μ is countably additive, as required. ■

Example 2.4. Let X be any countable set, say $X = \{\, x_1, x_2, \dots \}$ and let Σ be the collection of all subsets of X. Let (p_n) be any sequence of non-negative real numbers with $\sum_n p_n$ finite. If we define $\mu(A)$ for any $A \in \Sigma$ by

$$\mu(A) = \sum_{n \in I} p_n,$$

where $I = \{\, i : x_i \in A \,\}$, then μ is a finite measure on (X, Σ).

If $\sum_n p_n = 1$, then (X, Σ, μ) is a probability space. Such spaces are called discrete probability spaces.

Suppose that (X, Σ, μ) is a finite measure space and suppose that $A \in \Sigma$ with $\mu(A) = 0$. Suppose that $C \subset A$. If we think of μ as a probability or a volume, then $\mu(A) = 0$ should force $\mu(C) = 0$. Indeed, we might invoke Proposition 2.2 (part (iii)) to argue that

$$0 \le \mu(C) \le \mu(A) = 0$$

to conclude that, indeed, $\mu(C) = 0$. However, this argument is only valid if $C \in \Sigma$. Indeed, if $C \notin \Sigma$ then $\mu(C)$ is *not defined*. We can either accept this rather intuitively weird situation or we can endeavour to ensure that somehow $\mu(C)$ is defined in these circumstances and is equal to zero. This process is called "completing the measure". One proceeds as follows.

Let Σ' denote the collection of subsets of X satisfying $E \in \Sigma'$ if and only if there exist sets $A, B \in \Sigma$ with $A \subseteq E \subseteq B$ and $\mu(B \setminus A) = 0$ (so that $\mu(A) = \mu(B)$). Evidently, (take $A = B \in \Sigma$) we have $\Sigma \subseteq \Sigma'$.

Proposition 2.5. *Σ' is a σ-algebra.*

Proof. (i) $X \in \Sigma \subseteq \Sigma'$.

(ii) If $E \in \Sigma'$, then there is $A, B \in \Sigma$ with $A \subseteq E \subseteq B$ and $\mu(B \setminus A) = 0$. It follows that $B^c \subseteq E^c \subseteq A^c$ and $\mu(B^c) = \mu(X) - \mu(B) = \mu(X) - \mu(A) = \mu(A^c)$ and so $\mu(A^c \setminus B^c) = 0$. Hence $E^c \in \Sigma'$.

(iii) Let $E_1, E_2, \dots \in \Sigma'$. For each i, there is some $A_i, B_i \in \Sigma$ such that $A_i \subseteq E_i \subseteq B_i$ and $\mu(A_i) = \mu(B_i)$. It follows that $\bigcup_i A_i \subseteq \bigcup_i E_i \subseteq \bigcup_i B_i$. However, $\big(\bigcup_i B_i\big) \setminus \big(\bigcup_i A_i\big) \subseteq \bigcup_i (B_i \setminus A_i)$ and so

$$\mu\big(\,\big(\bigcup_i B_i\,\big) \setminus \big(\bigcup_i A_i\,\big)\,\big) \le \mu\big(\bigcup_i (B_i \setminus A_i)\,\big).$$

But $\bigcup_{i=1}^n (B_i \setminus A_i) \le \sum_{i=1}^n \mu(B_i \setminus A_i) = 0$ and $\bigcup_{i=1}^n (B_i \setminus A_i)$ increases to $\bigcup_{i=1}^\infty (B_i \setminus A_i)$ so that $\mu(\bigcup_{i=1}^n (B_i \setminus A_i)) \uparrow \mu(\bigcup_{i=1}^\infty (B_i \setminus A_i))$ which implies that $\mu(\bigcup_{i=1}^\infty (B_i \setminus A_i)) = 0$. Hence $\mu(\, (\bigcup_i B_i\,) \setminus (\bigcup_i A_i\,)\,) = 0$ and we finally conclude that $\bigcup_i E_i \in \Sigma'$. ∎

We wish to define the "measure" of an element of Σ'. Let $E \in \Sigma'$ and let $A, B \in \Sigma$ with $A \subseteq E \subseteq B$ and $\mu(A) = \mu(B)$. Since E is sandwiched between A and B and these have the same measure, it is clear what we must do, namely, we set $\mu'(E) = \mu(A)$. We must check that this value does not depend on any particular choice of the pair A, B. So suppose also that $A_1, B_1 \in \Sigma$ with $A_1 \subseteq E \subseteq B_1$ and $\mu(A_1) = \mu(B_1)$. Then

$$A \subseteq A \cup A_1 \subseteq E \subseteq B \cap B_1 \subseteq B$$

and therefore

$$\mu(A) \le \mu(A \cup A_1) \le \mu(B \cap B_1) \le \mu(B).$$

But $\mu(A) = \mu(B)$ and so we have equality throughout. In particular, we have $\mu(A) = \mu(A \cup A_1)$. Interchanging the pairs A, B and A_1, B_1, we similarly see that $\mu(A_1) = \mu(A_1 \cup A) = \mu(A)$ and so, in fact, $\mu(A) = \mu(B) = \mu(A_1) = \mu(B_1)$. It follows that the construction of μ' does not depend on any particular choices of A and B obeying the required conditions above.

Proposition 2.6. *μ' is an extension of μ on Σ to a measure on Σ', i.e., μ' is a measure on Σ' and $\mu'(A) = \mu(A)$ for all $A \in \Sigma$.*

Proof. For any $E \in \Sigma$, we may define $\mu'(E)$ by $\mu'(E) = \mu(E)$, since in this case we may simply take $A = E$ and $B = E$ to get the requirements $A \subseteq E \subseteq B$ with $\mu(A) = \mu(B)$. So trivially, μ' is an extension of μ.

We must show that μ' is a measure on Σ'. Evidently, $\mu'(E) \ge 0$ for any $E \in \Sigma'$.

To show countable additivity, suppose that $E_1, E_2, \dots \in \Sigma'$ are pairwise disjoint. For each $i \in \mathbb{N}$, there is A_i and B_i in Σ such that $A_i \subseteq E_i \subseteq B_i$ and $\mu(A_i) = \mu(B_i)$. Hence

$$\bigcup_i A_i \subseteq \bigcup_i E_i \subseteq \bigcup_i B_i$$

and we have seen in the proof of Proposition 2.5 that $\mu\big(\bigcup_i A_i\big) = \mu\big(\bigcup_i B_i\big)$. It follows from that the definition of μ' that $\mu'\big(\bigcup_i E_i\big) = \mu\big(\bigcup_i A_i\big)$. However, the A_is are pairwise disjoint (because $A_i \subseteq E_i$ and the E_is are) and so

$$\mu'\big(\bigcup_i E_i\big) = \mu\big(\bigcup_i A_i\big) = \sum_{i=1}^\infty \mu(A_i) = \sum_{i=1}^\infty \mu'(E_i)$$

since $\mu'(E_i) = \mu(A_i)$, which completes the proof. ∎

Definition 2.7. The measure space (X, Σ', μ') is called the completion of the space (X, Σ, μ).

Remark 2.8. Suppose that $A \in \Sigma$ and that $\mu(A) = 0$. Then for any subset $E \subseteq A$, we have $\varnothing \subseteq E \subseteq A$, with $\varnothing, A \in \Sigma$ and $\mu(\varnothing) = 0 = \mu(A)$. This means that $E \in \Sigma'$ and $\mu'(E) = 0$. So we can always complete a measure space to ensure that subsets of sets of measure zero also have "measure" zero, i.e., we consider the extension (X, Σ', μ') rather than (X, Σ, μ).

Definition 2.9. The measure space (X, Σ, μ) is called complete if $E \subseteq A$ with $A \in \Sigma$ and $\mu(A) = 0$ implies that $E \in \Sigma$ (and therefore $\mu(E) = 0$).

Remark 2.10. One sees that the completion of a measure space is complete and also that a compete measure space is equal to its completion (— if we try to complete a measure space that is already complete, then we get nothing new).

Chapter 3

Probability spaces, random variables and distribution functions.

Let $(\Omega, \mathcal{S}, \mathbb{P})$ be a probability space: so Ω is the sample space, $\mathcal{S}$ is the σ-algebra of events and $\mathbb{P}$ is the probability measure on $(\Omega, \mathcal{S})$. Let f be a random variable, i.e., $f : \Omega \to \mathbb{R}$ is Borel measurable. Hence, for any Borel set $A \subseteq \mathbb{R}$, the set $\{\, \omega \in \Omega : f(\omega) \in A \,\}$ belongs to $\mathcal{S}$, which means that it is an event and so its probability is defined. In other words, we can ask the question "what is the probability that f has its value in the (Borel) set A?" The answer is the value

$$\mathbb{P}(f^{-1}(A)) = \mathbb{P}(\{\, \omega \in \Omega : f(\omega) \in A \,\}).$$

We shall sometimes abbreviate this to just $\mathbb{P}(f \in A)$.

Definition 3.1. The distribution function of the random variable (RV) f is the function $F_f : \mathbb{R} \to \mathbb{R}$ given by

$$F_f(x) = \mathbb{P}(f \le x) \equiv \mathbb{P}(\{\, \omega : f(\omega) \le x \,\}).$$

Note that F_f is well-defined since $\{\, \omega : f(\omega) \le x \,\} \in \mathcal{S}$ for all $x \in \mathbb{R}$.

Proposition 3.2. *The distribution function F_f has the following properties.*

(i) $0 \le F_f(x) \le 1$ *for all* $x \in \mathbb{R}$.

(ii) $F_f(x) \le F_f(y)$ *whenever* $x \le y$.

(iii) $\lim_{x \to -\infty} F_f(x) = 0$ *and* $\lim_{x \to \infty} F_f(x) = 1$.

(iv) F_f *is continuous from the right, that is, for each* $x \in \mathbb{R}$, *we have* $F_f(x) = \lim_{h \downarrow 0} F_f(x + h)$.

Proof. (i) $F_f(x) = \mathbb{P}(f \le x) \in [0, 1]$ for all x.

(ii) For any $x \le y$, we have $\{\,\omega : f(\omega) \le x\,\} \subseteq \{\,\omega : f(\omega) \le y\,\}$ and so

$$F_f(x) = \mathbb{P}(\{\,\omega : f(\omega) \le x\,\}) \le \mathbb{P}(\{\,\omega : f(\omega) \le y\,\}) = F_f(y).$$

(iii) For $n \in \mathbb{N}$, let $E_n = \{\,\omega : f(\omega) \le -n\,\}$. Then $E_1 \supseteq E_2 \supseteq \dots$ and $\bigcap_n E_n = \varnothing$. Hence, by Proposition 2.2,

$$F_f(-n) = \mathbb{P}(E_n) \to \mathbb{P}(\varnothing) = 0$$

as $n \to \infty$. Given $\varepsilon > 0$, let n_0 be such that $\mathbb{P}(E_{n_0}) < \varepsilon$. Then for all $x < -n_0$, we have

$$\begin{aligned} 0 \le F_f(x) &\le F_f(-n_0), \quad \text{by (ii)}, \\ &= \mathbb{P}(E_{n_0}) \\ &< \varepsilon \end{aligned}$$

and the result follows.

Now let $A_n = \{\,\omega : f(\omega) \le n\,\}$ for $n \in \mathbb{N}$. Evidently, $A_1 \subseteq A_2 \subseteq \dots$ and $\bigcup_n A_n = \Omega$. Hence $\mathbb{P}(A_n) \to \mathbb{P}(\Omega) = 1$ as $n \to \infty$. Let $\varepsilon > 0$ be given and let n_0 be such that $\mathbb{P}(A_{n_0}) > 1 - \varepsilon$. Then, for all $x > n_0$, we have

$$\begin{aligned} 1 \ge F_f(x) &\ge F_f(n_0), \quad \text{by (ii)}, \\ &= \mathbb{P}(A_{n_0}) \\ &> 1 - \varepsilon \end{aligned}$$

and the result follows.

(iv) Fix $x \in \mathbb{R}$ and for $n \in \mathbb{N}$ set $B_n = \{\,\omega : f(\omega) \le x + \frac{1}{n}\,\}$. Then $B_1 \supseteq B_2 \supseteq \dots$ and $\bigcap_n B_n = \{\,\omega : f(\omega) \le x\,\}$. Again, by Proposition 2.2,

$$\mathbb{P}(B_n) \to \mathbb{P}(\{\,\omega : f(\omega) \le x\,\}) = F_f(x).$$

Let $\varepsilon > 0$ be given and let n_0 be such that $|\mathbb{P}(B_{n_0}) - F_f(x)| < \varepsilon$. Then, by part (ii),

$$0 \le \mathbb{P}(B_{n_0}) - F_f(x) < \varepsilon$$

i.e., $0 \le F_f\big(x + \frac{1}{n_0}\big) - F_f(x) < \varepsilon$. Let $0 < h < \frac{1}{n_0}$. Then, again by (ii),

$$\begin{aligned} 0 &\le F_f(x+h) - F_f(x) \\ &\le F_f\big(x + \frac{1}{n_0}\big) - F_f(x) \\ &< \varepsilon\,. \end{aligned}$$

Hence result. ■

Remark 3.3. F_f is an increasing function on $\mathbb{R}$ (i.e., is non-decreasing) and is bounded (by 1). Hence, if for given $a \in \mathbb{R}$ we have $x \uparrow a$, it follows that $F_f(x)$ increases to some limiting value — namely, $\sup_{x<a} F_f(x)$. Since F_f is increasing, this supremum is not greater than $F_f(a)$. In other words, $F_f(a)$ is an upper bound for the set $\{ F_f(x) : x < a \}$ and so is greater than (or equal to) the supremum of this set. Thus, the function F_f possesses a left-limit at each point in $\mathbb{R}$, but this limit value may be smaller than the actual value of F_f at that point. Denote by $F_f(a-)$ the limit of $F_f(x)$ as x increases to a. Then $F_f(a-) \le F_f(a)$.

Definition 3.4. The jump of F_f at $a \in \mathbb{R}$ is the difference $F_f(a) - F_f(a-)$. We say that a is a point of continuity of F_f if F_f is continuous at a, in which case $F_f(a) = F_f(a-)$ and so the jump is zero.

Proposition 3.5. *For any $a \in \mathbb{R}$, the jump of F_f at a is equal to $\mathbb{P}(f = a)$.*

Proof. By definition, $F_f(a) = \mathbb{P}(f \le a)$. Set $A_n = \{ \omega : f(\omega) \le a - \frac{1}{n} \}$ for $n \in \mathbb{N}$. Clearly, $A_1 \subseteq A_2 \subseteq \dots$ and $\bigcup_n A_n = \{ \omega : f(\omega) < a \}$. It follows that $\mathbb{P}(A_n) \uparrow \mathbb{P}(f < a)$. But

$$\lim_n \mathbb{P}(A_n) = \lim_n F_f(a - \tfrac{1}{n}) = F_f(a-)$$

so that $F_f(a-) = \mathbb{P}(f < a)$. Therefore the jump of F_f at a is

$$\text{“jump at } a\text{”} = F_f(a) - F_f(a-) = \mathbb{P}(f \le a) - \mathbb{P}(f < a) = \mathbb{P}(f = a),$$

as required. ∎

Let us say that a set is countable if it is either finite (including empty) or countably infinite (i.e., can be put into a one-one correspondence with the set of natural numbers $\mathbb{N}$).

Proposition 3.6. *The non-zero jumps of the distribution function F_f form a countable set.*

Proof. Denote by J the set of non-zero jumps of F_f and let J_n be the subset

$$J_n = \{ a \in J : \text{ “jump of } F_f \text{ at } a\text{”} \ge \tfrac{1}{n} \}.$$

Suppose $a_1, \dots a_k \in J_n$ with $a_1 < \dots < a_k$. Let a_0 be any real number such that $a_0 < a_1$. Then $0 \le F_f(a_0) \le F_f(a_k) \le 1$ and $F_f(a_i) - F_f(a_{i-1}) \ge \frac{1}{n}$ so that

$$1 \ge F_f(a_k) - F_f(a_0) = \sum_{i=1}^{k} (F_f(a_i) - F_f(a_{i-1})) \ge \sum_{i=1}^{k} \tfrac{1}{n} = \tfrac{k}{n}\,.$$

It follows that k cannot be larger than n, that is, J_n is either empty or otherwise cannot contain more than n elements. However, $J = \bigcup_n J_n$ and so J is a countable union of (empty or) finite sets and so is countable. ∎

Corollary 3.7. *For any random variable f, there is a countable set $J \subset \mathbb{R}$ such that $\mathbb{P}(f = x) = 0$ for all $x \in \mathbb{R} \setminus J$.*

Proof. Given f, let $J \subset \mathbb{R}$ denote the set of non-zero jumps of F_f. We have seen that J is countable. But $\mathbb{P}(f = x) \neq 0$ if and only if $x \in J$. ∎

Remark 3.8. The distribution function F_f of the random variable f is a non-decreasing function with values in the range $[0, 1]$ and it is continuous except possibly for countably-many jumps. Nonetheless, the behaviour of F_f can be quite complicated, as the illustrated by the following examples.

Examples 3.9.

1. Let $\Omega = \mathbb{N}$, $\mathcal{S}$ = all subsets of $\mathbb{N}$, and define $\mathbb{P}$ on $\Omega = \mathbb{N}$ by the assignment $\mathbb{P}(\{k\}) = 1/2^k$, for $k = 1, 2, \dots$. Then for any $A \subseteq \mathcal{S}$, $P(A) = \sum_{k \in A} \mathbb{P}(\{k\}) = \sum_{k \in A} 1/2^k$.

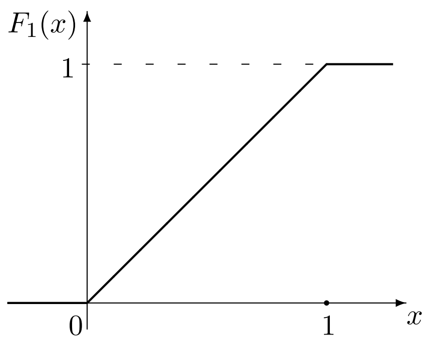
   Define f on $\mathbb{N}$ by $f(2k) = 1 + \frac{1}{k}$ and $f(2k+1) = -4 + \frac{1}{k}$. Then F_f has non-zero jumps at the points $2, 1 + \frac{1}{2}, 1 + \frac{1}{3}, 1 + \frac{1}{4}, \dots$, and also at $-3, -4 + \frac{1}{2}, -4 + \frac{1}{3}, -4 + \frac{1}{4}, \dots$.

   Notice that the jumps cluster at 1 and at -4. There are infinitely-many (but countably-many) jumps in any interval $(1, 1+\delta)$ and also in $(-4, -4+\delta)$, any $\delta > 0$.

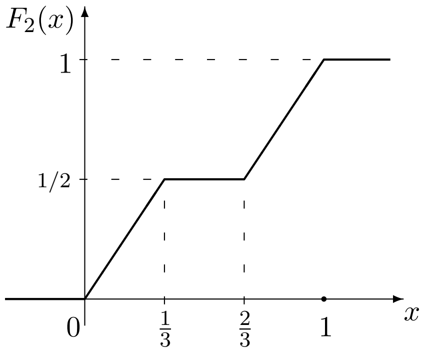
2. We can put more frills on this example as follows. Let $(\Omega, \mathcal{S}, \mathbb{P})$ be as above and let $r_1, r_2, \dots$ be an enumeration of the rationals, $\mathbb{Q}$. Define the random variable f on $\Omega = \mathbb{N}$ by $f(k) = r_k$, $k \in \mathbb{N}$. Then F_f has a non-zero jump at *every* rational point in $\mathbb{R}$. The value of the jump of F_f at the rational point r_k is $\mathbb{P}(f = r_k) = \mathbb{P}(\{k\}) = 1/2^k$. Note that the function F_f is *continuous* at every irrational in $\mathbb{R}$.

Example 3.10 (Cantor). We shall construct a particular non-decreasing function $F(x)$ on $\mathbb{R}$, obeying $0 \leq F(x) \leq 1$, by means of a certain limiting procedure. We start with $F_1(x)$ defined as follows:

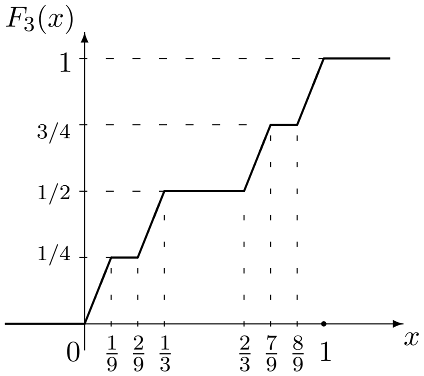
$$F_1(x) = \begin{cases} 0, & x \leq 0, \\ x, & 0 \leq x \leq 1, \\ 1, & x \geq 1. \end{cases}$$



Next, we construct F_2 from F_1 by "flattening out" the middle third.



F_3 is then constructed from F_2 by "flattening out" the middle thirds over $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$.



F_4 is obtained from F_3 by "flattening out" the portions over the intervals $[0, \frac{1}{9}]$, $[\frac{2}{9}, \frac{1}{3}]$, $[\frac{2}{3}, \frac{7}{9}]$ and $[\frac{8}{9}, 1]$ ...and so on.

In this way, we obtain a sequence of functions, (F_n). We make the following observations.

(i) Each $F_n(x)$ is non-decreasing, continuous, is zero for all negative values of x and is 1 for all values of $x \geq 1$.

(ii) The "flat" parts of F_{n+1} contain the "flat" parts of F_n (so that F_{n+1} has more flat parts than F_n).

(iii) For any $x \in \mathbb{R}$, $|F_{n+1}(x) - F_n(x)| \leq 1/2^n$.

We will show that $(F_n(x))$ is a Cauchy sequence (uniformly in x). Indeed, for $n > m$, we have

$$\begin{aligned}
&|F_n(x) - F_m(x)| \\
&\qquad = |F_n(x) - F_{n-1}(x) + F_{n-1}(x) - F_{n-2}(x) + \cdots + F_{m+1}(x) - F_m(x)| \\
&\qquad \leq |F_n(x) - F_{n-1}(x)| + |F_{n-1}(x) - F_{n-2}(x)| + \dots \\
&\qquad\qquad\qquad\qquad\qquad\qquad \cdots + |F_{m+1}(x) - F_m(x)| \\
&\qquad \leq \frac{1}{2^{n-1}} + \frac{1}{2^{n-2}} + \cdots + \frac{1}{2^m} \\
&\qquad = \frac{1}{2^{m-1}}\Big(\frac{1}{2^{n-m}} + \cdots + \frac{1}{2}\Big) \\
&\qquad < \frac{1}{2^{m-1}}, \quad \text{since } \textstyle\sum_{k=1}^{\infty} \frac{1}{2^k} = 1.
\end{aligned}$$

It follows that $(F_n(x))$ is a Cauchy sequence, uniformly in x, as claimed. Hence $F_n(x)$ converges uniformly in $\mathbb{R}$, as $n \to \infty$. Call the limit $F(x)$.
(a) Evidently, $F(x) = 0$ for all $x \leq 0$ and $F(x) = 1$ for all $x \geq 1$ (because these properties hold for every F_n). Furthermore, if $x \leq y$, then we have $F_n(x) \leq F_n(y)$, i.e., $F_n(x) - F_n(y) \leq 0$ for all n. Taking the limit $n \to \infty$, gives $F(x) - F(y) \leq 0$, i.e., $F(x) \leq F(y)$. So F is non-decreasing.

(b) $F(x)$ is continuous. This is a direct consequence of the uniform convergence together with the continuity of each $F_n(x)$.

Now, F is “flat” whenever any F_n is, by observation (ii) above. That is, F is differentiable, with $F' = 0$, on each of the intervals $I_1, I_2, \dots$ given as $(\frac{1}{3}, \frac{2}{3})$, $(\frac{1}{9}, \frac{2}{9})$, $(\frac{7}{9}, \frac{8}{9})$, $(\frac{1}{27}, \frac{2}{27})$, $(\frac{4}{27}, \frac{5}{27})$, $(\frac{22}{27}, \frac{23}{27})$, $(\frac{25}{27}, \frac{26}{27})$, $\dots$ in $[0, 1]$.

The total length of these intervals is

$$\begin{aligned}
\frac{1}{3} + \frac{2}{3^2} + \frac{4}{3^3} + \frac{8}{3^4} + \frac{16}{3^5} + \dots \quad &= \frac{1}{3} + \frac{2}{3^2} + \frac{2^2}{3^3} + \frac{2^3}{3^4} + \frac{2^4}{3^5} + \dots \\
&= \frac{1}{3}\Big[1 + \frac{2}{3} + \Big(\frac{2}{3}\Big)^2 + \Big(\frac{2}{3}\Big)^3 + \dots\Big] \\
&= \frac{1}{3}\,\frac{1}{(1 - 2/3)} = 1.
\end{aligned}$$

Thus, $F'(x) = 0$ for $x \in I_1 \cup I_2 \cup \dots$ (and also for all $x < 0$, and all $x > 0$). The total length of $I_1 \cup I_2 \cup \cdots \subset [0, 1]$ is equal to 1. But nonetheless, F is continuous and increases from 0 to 1 over the interval $[0, 1]$.

Remark 3.11. The above example was of a function F satisfying the properties of a distribution function as given by Proposition 3.2. However, it is not yet clear that this F actually *is* a distribution function corresponding to some random variable on some probability space. We will see later that this is indeed the case.

Remark 3.12. Let (X, Σ, μ) be a finite measure space and let $f : X \to \mathbb{R}$ be Borel measurable. For any Borel set $E \subseteq \mathbb{R}$, define $\nu(E) = \mu(f^{-1}(E))$, that is,

$$\nu(E) = \mu(\{ x \in X : f(x) \in E \}).$$

Then it is straightforward to check that ν defines a finite measure on $\mathcal{B}(\mathbb{R})$. In particular, if f is a random variable on $(\Omega, \mathcal{S}, \mathbb{P})$, then the assignment $\nu(E) = \mathbb{P}(\{ \omega \in \Omega : f(\omega) \in E \})$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Moreover,

$$\nu((-\infty, a]) = \mathbb{P}(f \le a) = F_f(a).$$

The sets of the form $\{ (-\infty, a] : a \in \mathbb{R} \}$ generate $\mathcal{B}(\mathbb{R})$ and so ν is completely determined by the values $\{ \nu((-\infty, a]) : a \in \mathbb{R} \}$, that is, ν is determined by F_f.

(Suppose that μ_1 and μ_2 are finite measures on a measurable space (X, Σ) such that $\mu_1(X) = \mu_2(X)$. Then the collection $S = \{ E \in \Sigma : \mu_1(E) = \mu_2(E) \}$ is a sub-σ-algebra of Σ. If S contains a set which generates Σ then it follows that $S = \Sigma$.)

Suppose now that ν is a given probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Set $F(x) = \nu((-\infty, x])$ for $x \in \mathbb{R}$. Then using the properties of ν, we see that F has values in $[0, 1]$, is non-decreasing, is continuous from the right etc, i.e., F satisfies all properties of Proposition 3.2.

Two questions now spring to mind.

1. Given a function F satisfying the properties of Proposition 3.2, is there a probability measure ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $F(x) = \nu((-\infty, x])$ for all $x \in \mathbb{R}$?

2. Given some function F satisfying the properties of Proposition 3.2, is F the distribution function of some random variable?

The answer to both these questions is "yes". The first is dealt with by the following theorem.

Theorem 3.13. *For any given function F satisfying the properties of Proposition 3.2, there exists a probability measure ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfying $F(x) = \nu((-\infty, x])$ for all $x \in \mathbb{R}$. Moreover, such ν is unique.*

We will not prove this here, but we note that as a consequence,

$$\begin{aligned}
\nu((-\infty, a)) &= F(a-), \\
\nu((a, b]) &= F(b) - F(a+), \quad a < b, \\
\nu((a, b)) &= F(b-) - F(a+), \quad a < b, \\
\nu([a, b)) &= F(b-) - F(a-), \quad a < b, \\
\nu([a, b]) &= F(b) - F(a-), \quad a \le b.
\end{aligned}$$

The measure associated with F in this way is called the Lebesgue-Stieltjes measure given by F. In particular, suppose that F is given by

$$F(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \le x \le 1, \\ 1, & x > 1. \end{cases}$$

Then the completed measure given by F, when thought of as a measure on $[0, 1]$, is called Lebesgue measure on $[0, 1]$. For this F, we have $\nu((-\infty, 0]) = F(0) = 0$ and $\nu((1, \infty)) = 1 - F(1) = 0$. For $0 \le a \le b \le 1$, we have

$$\begin{aligned} \nu((a, b)) &= \nu([a, b)) = \nu((a, b]) \\ &= \nu([a, b]) \\ &= b - a \end{aligned}$$

so that ν is just "length".

Remark 3.14. If we replace $\lim_{x\to\infty} F(x) = 1$ by, say, $\lim_{x\to\infty} F(x) = M$, then $\nu(\mathbb{R}) = M$ rather than 1. We could also consider the function

$$F(x) = \begin{cases} 0, & x < a, \\ x - a, & a \le x \le b, \\ b - a, & x > b. \end{cases}$$

This would then give Lebesgue measure on $[a, b]$ — it still corresponds to "length".

We can now answer the second question in the affirmative.

Theorem 3.15. *Suppose that F satisfies the properties of Proposition 3.2. Then there exists a random variable f such that $F = F_f$.*

Proof. We must first specify a probability space on which f will be defined. Take $\Omega = \mathbb{R}$, $\mathcal{S} = \mathcal{B}(\mathbb{R})$ and let $\mathbb{P}$ be the Lebesgue-Stieltjes probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ determined by F according to Theorem 3.13.

Define the random variable f on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ by $f(x) = x$. Then f *is* a random variable (i.e., it is Borel measurable) and

$$\begin{aligned} F_f(a) &= \mathrm{Prob}(f \le a) \\ &= \mathbb{P}(\{\, x \in R : f(x) \le a \,\}) \\ &= \mathbb{P}(\{\, x \in \mathbb{R} : x \le a \,\}) \\ &= \mathbb{P}((-\infty, a])) \\ &= F(a), \end{aligned}$$

by the construction of the Lebesgue-Stieltjes measure $\mathbb{P}$ from F. ∎

Remark 3.16. This result guarantees the existence of a random variable with any preassigned distribution. For example, let

$$F(x) = \int_{-\infty}^{x} e^{-t^2/2} \frac{dt}{\sqrt{2\pi}}\,.$$

Then by the Theorem, there *is* a random variable with this F as its distribution function. We conclude that standard normal random variables do exist!

Also, there is a random variable f such that its distribution function is the function F in the Cantor Example 3.10. Note that if λ denotes Lebesgue measure, then

$$\begin{aligned}\lambda(I_1 \cup I_2 \cup \dots) &= \lambda(I_1) + \lambda(I_2) + \dots \\ &= \text{“sum of lengths”} \\ &= 1.\end{aligned}$$

But if ν is the Lebesgue-Stieltjes measure given by F in this example, then we find that $\nu(I_1) = \nu(I_2) = \dots = 0$ since F is constant on each of the I_js (so that $\nu(I_j) = F(b_j-) - F(a_j) = 0$, where $I_j = (a_j, b_j)$). It follows that $\nu(I_1 \cup I_2 \cup \dots) = 0$. The measure ν is “singular” when compared with Lebesgue measure.

Chapter 4

Integration theory.

We will develop the theory of integration on an arbitrary finite measure space (X, Σ, μ). Recall that a simple function s is one which takes only finitely-many values, $\alpha_1, \dots, \alpha_n$, say, so that $s(x) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(x)$ for disjoint subsets $A_1, \dots, A_n$ in X. To avoid tedious repetition, we will use the word "simple" to mean "measurable and simple". Thus, the sets A_i belong to Σ.

Definition 4.1. Suppose that $s = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$ is a simple function. For given $E \in \Sigma$, we define the integral of s over E with respect to μ to be

$$\int_E s \, d\mu = \sum_{i=1}^n \alpha_i \, \mu(A_i \cap E).$$

Thus, in particular, for any $A \in \Sigma$, the integral of the indicator function of A over X is just the measure of A:

$$\int_X \mathbb{1}_A \, d\mu = \mu(A).$$

If μ is Lebesgue measure on $[0, 1]$ and $A = [a, b]$ with $0 \le a \le b \le 1$, then $\int_{[0,1]} \mathbb{1}_{[a,b]} \, d\mu = b - a$, which is just the usual Riemann integral of $\mathbb{1}_{[a,b]}$ over the interval $[0, 1]$.

If $X = \mathbb{R}$ and μ is the Lebesgue-Stieltjes measure on $\mathbb{R}$ as determined by the function $F(x) = \int_{-\infty}^x \rho(t) \, dt$, where ρ is a given non-negative Riemann-integrable function, then for $a < b$,

$$\int_{\mathbb{R}} \mathbb{1}_{[a,b]} \, d\mu = \mu([a, b]) = F(b) - F(a) = \int_a^b \rho(t) \, dt.$$

Definition 4.2. Let $f : X \to \mathbb{R}$ be measurable and suppose that $f \ge 0$. For any $E \in \Sigma$, we define

$$\int_E f \, d\mu = sup \int_E s \, d\mu$$

where the supremum is taken over all non-negative simple functions s satisfying $0 \le s(x) \le f(x)$, for $x \in X$.

If the right hand side is not finite, then we say that f is not integrable over E (or has "infinite integral" over E).

Note that there do exist simple functions s obeying $0 \le s(x) \le f(x)$ for all $x \in X$. The function $s(x) = 0 = 0\,\mathbb{1}_X(x)$ is one such. Various properties of the integral are collected next.

Proposition 4.3. *Suppose that f, g are measurable and let $E \in \Sigma$.*

(i) *If $0 \le f \le g$, then $\int_E f\,d\mu \le \int_E g\,d\mu$.*

(ii) *If $A \subseteq B$, $A, B \in \Sigma$, and $f \ge 0$, then $\int_A f\,d\mu \le \int_B f\,d\mu$.*

(iii) *If $f \ge 0$ and $c \ge 0$ is a constant, then $\int_E cf\,d\mu = c\int_E f\,d\mu$.*

(iv) *If $f(x) = 0$ for all $x \in E$, then $\int_E f\,d\mu = 0$.*

(v) *If $\mu(E) = 0$, then $\int_E f\,d\mu = 0$ for any $f \ge 0$.*

(vi) *If $f \ge 0$, then $\int_E f\,d\mu = \int_X \mathbb{1}_E f\,d\mu$.*

Proof. These are all fairly clear from the definitions.

(i) If s is simple and $0 \le s \le f$, then $0 \le s \le f \le g$ and so, by definition of $\int_E g\,d\mu$, we have $\int_E s\,d\mu \le \int_E g\,d\mu$. Taking the supremum over all s with $0 \le s \le f$ gives $\int_E f\,d\mu \le \int_E g\,d\mu$.

(ii) If $0 \le s \le f$, then $\int_A s\,d\mu \le \int_B s\,d\mu$ (since $A \subseteq B$) $\le \int_B f\,d\mu$. Taking the supremum over s with $0 \le s \le f$ gives the inequality $\int_A f\,d\mu \le \int_B f\,d\mu$, as required.

(iii) We may assume that $c \ne 0$. Then

$$\begin{aligned}\int_E c\,f\,d\mu &= \sup_{\{\,s\,:\,0\le s\le f\,\}} \int_E c\,s\,d\mu \\ &= c \sup_{\{\,s\,:\,0\le s\le f\,\}} \int_E s\,d\mu \\ &= c\int_E f\,d\mu.\end{aligned}$$

The first equality here takes a moment's thought. Any simple function s_1 with $0 \le s_1 \le cf$ has the form $s_1 = cs_2$ for simple $s_2 (= s_1/c)$ obeying $0 \le s_2 \le f$ and, conversely, every if s_2 satisfies $0 \le s_2 \le f$, then $s_1 = cs_2$ satisfies $0 \le s_1 \le cf$.

(iv) If $f(x) = 0$ for $x \in E$, then $0 \le s \le f$ forces $s = 0$ on E. Hence, if $A_i = s^{-1}(\alpha_i)$ and $A_i \cap E \ne \varnothing$, we must have $\alpha_i = 0$. If $A_i \cap E = \varnothing$ then $\mu(A_i \cap E) = 0$. It follows from the definition that $\int_E s\,d\mu = 0$. Hence the supremum over all such s must also be zero.

(v) If $\mu(E) = 0$, then $\int_E s\,d\mu = 0$ for any simple s.

(vi) Clearly, if $s \geq 0$ is simple, then so is $\mathbb{1}_E s$ and $\int_E s\, d\mu = \int_X \mathbb{1}_E\, s\, d\mu$. Now let $0 \leq s \leq f$. Then $0 \leq \mathbb{1}_E\, s \leq \mathbb{1}_E\, f$ and so

$$\int_E s\, d\mu = \int_X \mathbb{1}_E s\, d\mu \leq \int_X \mathbb{1}_E f\, d\mu.$$

Taking the supremum over such s gives

$$\int_E f\, d\mu \leq \int_X \mathbb{1}_E f\, d\mu.$$

On the other hand, suppose that $0 \leq s \leq \mathbb{1}_E f$. Then certainly $0 \leq s \leq f$ and s vanishes off E so $s = \mathbb{1}_E s$. Hence

$$\int_X s\, d\mu = \int_X \mathbb{1}_E s\, d\mu = \int_E s\, d\mu \leq \int_E f\, d\mu.$$

Taking the supremum gives

$$\int_X \mathbb{1}_E f\, d\mu \leq \int_E f\, d\mu$$

and the equality follows. ∎

Note that (i) and (vi) together imply (ii) and that (vi) implies (iv).

Proposition 4.4. *Let s, t be any given simple functions with $s \geq 0$ and $t \geq 0$. For $E \in \Sigma$, put $\varphi(E) = \int_E s\, d\mu$. Then φ is a finite measure on (X, Σ). Furthermore,*

$$\int_X (s+t)\, d\mu = \int_X s\, d\mu + \int_X t\, d\mu.$$

Proof. Suppose $s = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$. Then

$$\varphi(E) = \int_E s\, d\mu = \sum_{i=1}^n \alpha_i\, \mu(A_i \cap E).$$

Each map $E \mapsto \mu(A_i \cap E)$ is a finite measure on Σ and so therefore is their finite sum over $1 \leq i \leq n$.

To show that $\int_X (s+t)\, d\mu = \int_X s\, d\mu + \int_X t\, d\mu$, let us write $s = \sum_{i=0}^n \alpha_i \mathbb{1}_{A_i}$ and $t = \sum_{j=0}^m \beta_j \mathbb{1}_{B_j}$ where $\alpha_0 = 0 = \beta_0$, the α_is are distinct and the β_js are distinct so that $\{ A_0, \dots, A_n \}$ is a partition of X and similarly, $\{ B_0, \dots, B_m \}$ is a partition of X. The sum $s + t$ is a simple function, say, $s + t = \sum_{\ell=0}^k \gamma_\ell \mathbb{1}_{C_\ell}$, where $\gamma_0 = 0$, the γ_ℓs are distinct and $\{ C_0, \dots, C_k \}$ is a partition of X. Each γ_ℓ has the form $\gamma_\ell = \alpha_i + \beta_j$ for suitable i, j and so $C_\ell = \bigcup_{I_\ell} A_i \cap B_j$ where $I_\ell = \{ (i,j) : \alpha_i + \beta_j = \gamma_\ell \}$. Then

$$\int_X (s+t)\, d\mu = \sum_{\ell=0}^k \gamma_\ell\, \mu(C_\ell)$$

and

$$\int_X s\,d\mu + \int_X t\,d\mu = \sum_{i=0}^{n} \alpha_i\,\mu(A_i) + \sum_{j=0}^{m} \beta_j\,\mu(B_j).$$

The collection $\{\, A_i \cap B_j : 0 \le i \le n,\ 0 \le j \le m \,\}$ forms a partition of X. We have

$$\begin{aligned}
\int_X (s+t)\,d\mu &= \sum_{\ell=0}^{k} \gamma_\ell\,\mu(C_\ell) \\
&= \sum_{\ell=0}^{k} \gamma_\ell\,\mu\big(\bigcup_{I_\ell} A_i \cap B_j\big) \\
&= \sum_{\ell=0}^{k} \gamma_\ell \sum_{I_\ell} \mu(A_i \cap B_j), \quad \text{since distinct } A_i \cap B_j\text{s are disjoint,} \\
&= \sum_{\ell=0}^{k} \sum_{I_\ell} (\alpha_i + \beta_j)\,\mu(A_i \cap B_j) \\
&= \sum_{i,j} (\alpha_i + \beta_j)\,\mu(A_i \cap B_j) \\
&= \sum_{i,j} \alpha_i\,\mu(A_i \cap B_j) + \sum_{i,j} \beta_j\,\mu(A_i \cap B_j) \\
&= \sum_{i} \alpha_i\,\mu(A_i) + \sum_{j} \beta_j\,\mu(B_j) \\
&= \int_X s\,d\mu + \int_X t\,d\mu
\end{aligned}$$

and the proof is complete. ∎

Remark 4.5. From this result, we see that the integral of a simple function is got by adding up its "elementary bits" and it does not matter how we write the simple function as such a sum. For example, consider $s = \alpha \mathbb{1}_A$ so that $\int s\,d\mu = \alpha\mu(A)$. If $A = B_1 \cup B_2$ with $B_1 \cap B_2 = \varnothing$, then we could also write s as $s = \alpha \mathbb{1}_{B_1} + \alpha \mathbb{1}_{B_2}$. We have

$$\begin{aligned}
\int_X \alpha \mathbb{1}_A\,d\mu &= \int_X \alpha \mathbb{1}_{B_1}\,d\mu + \int_X \alpha \mathbb{1}_{B_2}\,d\mu \\
&= \alpha\mu(B_1) + \alpha\mu(B_2) \\
&= \alpha\mu(A)
\end{aligned}$$

as it should.

In general, if s is simple, we can write $s = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ and we do not have to worry whether the α_is are all distinct, or indeed whether the A_is are disjoint. In any event, $\int_X s\,d\mu$ is $\sum_{i=1}^{n} \alpha_i \mu(A_i)$.

The following theorem is crucial for further development.

Theorem 4.6 (Lebesgue's Monotone Convergence Theorem). *Let (f_n) be a sequence of measurable functions on X and suppose that*

(i) $0 \le f_1(x) \le f_2(x) \le \dots$ *for each* $x \in X$,

(ii) $f_n(x) \to f(x)$ *as* $n \to \infty$, *for each* $x \in X$.

Then f is measurable and $\int_X f_n \, d\mu \to \int_X f \, d\mu$ as $n \to \infty$.

(Note: this last statement means that if some $f_n$is not integrable, then neither is f, or if $\int_X f_n \, d\mu$ diverges then f is not integrable.)

Proof. The limit of a sequence of measurable functions is measurable and so f is measurable (Theorem 1.16). Also, since $0 \le f_1(x) \le f_2(x) \le \dots$ and $f_n(x) \to f(x)$, it is clear that $f_n(x) \le f(x)$ for all $n \in \mathbb{N}$ and all $x \in X$.

Suppose f_N is not integrable. Then for any $M > 0$ there is a simple function s with $0 \le s \le f_N$ such that $\int_X s \; dmu > M$. Since $0 \le s \le f$ we deduce that $\sup\{ \int_X s \, d\mu : 0 \le s \le f \}$ is infinite, i.e., f is not integrable.

Suppose now that each f_n is integrable. Since $f_n \le f_{n+1}$ it follows that $\int_X f_n \, d\mu \le \int_X f_{n+1} \, d\mu$. Hence the non-decreasing sequence $(\int_X f_n \, d\mu)$ (of non-negative numbers) either diverges, as $n \to \infty$, or else converges to some $\alpha \ge 0$ (its supremum).

Suppose first that the sequence diverges. Then for any $M > 0$, there is N such that $\int_X f_N \, d\mu > M + 1$. By definition of the integral, there is a simple function s obeying $0 \le s \le f_N$ such that $\int_X s \, d\mu > M$. But then $0 \le s \le f$ and it follows that $\sup\{ \int_X s \, d\mu : 0 \le s \le f \}$ is infinite and hence f is not integrable.

Now suppose that the sequence $\int_X f_n \, d\mu$ converges to $\alpha \ge 0$. We must show that f is integrable and that $\int_X f \, d\mu = \alpha$. Let s be any simple function such that $0 \le s \le f$ and let c be a constant with $0 < c < 1$. For each $n \in \mathbb{N}$, set

$$E_n = \{ x \in X : f_n(x) \ge c \, s(x) \} .$$

Then $E_n \in \Sigma$ and, because $(f_n(x))$ is increasing, $E_1 \subseteq E_2 \subseteq E_3 \subseteq \dots$. Moreover, since $f_n(x) \to f(x) \ge s(x) \ge c \, s(x)$, it follows that $\bigcup_n E_n = X$. (If $f(x) = 0$, then also $s(x) = 0$ and $x \in E_n$ for all n. On the other hand, if $f(x) \ne 0$, then $f(x) \ge s(x)$ implies that $f(x)/c > f(x) \ge s(x)$ and so $f(x) > c \, s(x)$. It follows that for each x with $f(x) \ne 0$, $f_n(x) > c \, s(x)$ for all sufficiently large n (which, of course, may depend on x). In other words, any such x belongs to E_n for sufficiently large n.)

By Proposition 4.3, we have

$$\int_X f_n \, d\mu \ge \int_{E_n} f_n \, d\mu \ge c \int_{E_n} s \, d\mu.$$

Now, $E \mapsto \int_E s\,d\mu$ is a measure and $E_n \uparrow X$ and so letting $n \to \infty$ we get

$$\alpha \geq c \int_X s\,d\mu.$$

This holds for any $0 < c < 1$ and so we must have $\alpha \geq \int_X s\,d\mu$. It follows that $\sup\{\int_X s\,d\mu : 0 \leq s \leq f\} \leq \alpha$, which shows that f is integrable and also that $\int_X f\,d\mu \leq \alpha$.

On the other hand, $f_n \leq f$ and so $\int_X f_n\,d\mu \leq \int_X f\,d\mu$ for all n. Hence, letting $n \to \infty$, we see that $\alpha \leq \int_X f\,d\mu$. We conclude that, in fact, $\alpha = \int_X f\,d\mu$, that is, $\lim_n \int_X f_n\,d\mu = \int_X f\,d\mu$. ■

Corollary 4.7. *Suppose that $f \geq 0$ and $g \geq 0$ and that both f and g are integrable. Then $f + g$ is integrable and*

$$\int_X (f+g)\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu.$$

Proof. Let (s_n) and (t_n) be increasing sequences of non-negative simple functions such that $s_n(x) \to f(x)$ and $t_n(x) \to g(x)$ for $x \in X$. Then $(s_n + t_n)$ is an increasing sequence of non-negative simple functions with $(s_n + t_n)(x) \to (f+g)(x)$ for each $x \in X$. Now

$$\int_X (s_n + t_n)\,d\mu = \int_X s_n\,d\mu + \int_X t_n\,d\mu \qquad (*)$$

By Lebesgue's Monotone Convergence Theorem, Theorem 4.6, the right hand side of $(*)$ converges to $\int_X f\,d\mu + \int_X g\,d\mu$. Therefore the left hand side must also converge, and so again by Lebesgue's Monotone Convergence Theorem, we deduce that $f + g$ is integrable and that

$$\int_X (f+g)\,d\mu = \lim_n \int_X (s_n + t_n)\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu$$

as required. ■

Corollary 4.8. *Suppose that $f \geq 0$ is integrable. Then the set function $E \mapsto \int_E f\,d\mu$ is a finite measure on (X, Σ).*

Proof. Let $\varphi(E) = \int_E f\,d\mu = \int_X \mathbb{1}_E\, f\,d\mu$. By Corollary 4.7, it follows that $E \mapsto \varphi(E)$ is finitely-additive, that is, $\varphi(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n \varphi(E_i)$ for any finite set $E_1, \dots, E_n$ of pairwise disjoint sets in Σ. We must show that φ is, in fact, σ-additive. To show this, suppose that (E_n) is any sequence in Σ with $E_n \supseteq E_{n+1}$ and $\bigcap_n E_n = \varnothing$. We claim that $\varphi(E_n) \to 0$. Indeed, $E_n \downarrow \varnothing$ and so $1 - \mathbb{1}_{E_n} \uparrow 1$ as $n \to \infty$. Hence $(1 - \mathbb{1}_{E_n})\, f \uparrow f$ and so, by Lebesgue's Monotone Convergence Theorem, $\int_X (1 - \mathbb{1}_{E_n})\, f\,d\mu \uparrow \int_X f\,d\mu$. In other words,

$$\int_X f\,d\mu - \int_X \mathbb{1}_{E_n} f\,d\mu \uparrow \int_X f\,d\mu$$

and so $\varphi(E_n) \downarrow 0$ as claimed and therefore φ is σ-additive. ■

So far, we have only considered the integrability of non-negative functions. The general situation is handled by linearity as follows.

Definition 4.9. The complex-valued function f on X is said to be (Lebesgue) integrable (with respect to μ) if $|f|$ is integrable. The collection of all such functions is denoted by $\mathcal{L}^1(X,\mu)$.

For any $f = u + iv \in \mathcal{L}^1(X,\mu)$, we define

$$\int_X f\,d\mu = \int_X u_+\,d\mu - \int_X u_-\,d\mu + i\int_X v_+\,d\mu - i\int_X v_-\,d\mu$$

where $u_\pm$ and $v_\pm$ are the positive and negative parts of u and v, respectively. (For any real-valued function g, $g_\pm = \frac{1}{2}(|g| \pm g)$ so that both $g_\pm \geq 0$, $g_+g_- = 0$ and $g = g_+ - g_-$ and $|g| = g_+ + g_-$.)

Note that $u_\pm \geq 0$ on X and $u_\pm \leq |u| \leq |f|$ and so $u_\pm$ is (measurable and) integrable on X with the similar remark applying to $v_\pm$. It follows that $\int_X f\,d\mu$ is well-defined.

Theorem 4.10. *If $f, g \in \mathcal{L}^1(X,\mu)$ and $a, b \in \mathbb{C}$, then $af + bg \in \mathcal{L}^1(X,\mu)$ and*

$$\int_X (af + bg)\,d\mu = a\int_X f\,d\mu + b\int_X g\,d\mu.$$

Proof. We first note that $|af + bg| \leq |a|\,|f| + |b|\,|g|$ and so $af + bg \in \mathcal{L}^1(X,\mu)$ and

$$\int_X |af + bg|\;d\mu \leq |a|\int_X |f|\;d\mu + |b|\int_X |g|\;d\mu.$$

It is enough to prove that $\int_X (f+g)\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu$ and that $\int_X af\,d\mu = a\int_X f\,d\mu$.

Suppose that f and g are real-valued. Let $h = f + g$. Then

$$h = h_+ - h_- = f_+ - f_- + g_+ - g_-$$

and so $h_+ + f_- + g_- = h_- + f_+ + g_+$ which implies that $\int_X (h_+ + f_- + g_-)\,d\mu = \int_X (h_- + f_+ + g_+)\,d\mu$. But then, by Corollary 4.7, we have

$$\int_X h_+\,d\mu + \int_X f_-\,d\mu + \int_X g_-\,d\mu = \int_X h_-\,d\mu + \int_X f_+\,d\mu + \int_X g_+\,d\mu$$

which becomes

$$\int_X h_+\,d\mu - \int_X h_-\,d\mu = \int_X f_+\,d\mu - \int_X f_-\,d\mu + \int_X g_+\,d\mu - \int_X g_-\,d\mu\,,$$

that is, by definition,

$$\int_X h\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu.$$

The case of complex-valued f and g now follows by looking at the real and imaginary parts. Indeed, if $h = f + g$, then, by definition, $\int_X h\,d\mu = \int_X \operatorname{Re} h\,d\mu + i\int_X \operatorname{Im} h\,d\mu$, $\int_X f\,d\mu = \int_X \operatorname{Re} f\,d\mu + i\int_X \operatorname{Im} f\,d\mu$ and $\int_X g\,d\mu = \int_X \operatorname{Re} g\,d\mu + i\int_X \operatorname{Im} g\,d\mu$. But $\operatorname{Re} h = \operatorname{Re} f + \operatorname{Re} g$ and $\operatorname{Im} h = \operatorname{Im} f + \operatorname{Im} g$ and, from the argument above, $\int_X (\operatorname{Re} f + \operatorname{Re} g)\,d\mu = \int_X \operatorname{Re} f\,d\mu + \int_X \operatorname{Re} g\,d\mu$ and $\int_X (\operatorname{Im} f + \operatorname{Im} g)\,d\mu = \int_X \operatorname{Im} f\,d\mu + \int_X \operatorname{Im} g\,d\mu$. It follows that $\int_X h\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu$.

To show that $\int_X af\,d\mu = a\int_X f\,d\mu$ for all $a \in \mathbb{C}$, it is enough to consider the cases (i) a real and positive, (ii) $a = -1$ and (iii) $a = i$. (The general case then follows using this together with the previous part.)
(i) Suppose $a \geq 0$. Then $af = au_+ - au_- + iav_+ - iav_-$ and from the above,

$$\begin{aligned}\int_X af\,d\mu &= \int_X au_+\,d\mu - \int_X au_-\,d\mu + i\int_X av_+\,d\mu - i\int_X av_-\,d\mu \\ &= a\int_X u_+\,d\mu - a\int_X u_-\,d\mu + ia\int_X v_+\,d\mu - ia\int_X v_-\,d\mu \\ &= a\int_X f\,d\mu.\end{aligned}$$

(ii) Let $a = -1$. Then

$$\begin{aligned}\int_X -f\,d\mu &= \int_X u_-\,d\mu - \int_X u_+\,d\mu + i\int_X v_-\,d\mu - i\int_X v_+\,d\mu \\ &= -\int_X f\,d\mu.\end{aligned}$$

(iii) Finally, let $a = i$. Then

$$\begin{aligned}\int_X if\,d\mu &= \int_X (i(u_+ - u_-) + (v_- - v_+))\,d\mu \\ &= \int_X v_-\,d\mu - \int_X v_+\,d\mu + i\int_X u_+\,d\mu - i\int_X u_-\,d\mu \\ &= i\int_X f\,d\mu.\end{aligned}$$

Using all the above results yields $\int_X af\,d\mu = a\int_X f\,d\mu$ for $a \in \mathbb{C}$. ■

Remark 4.11. The above result says that the space $\mathcal{L}^1(X,\mu)$ is a (complex) linear space.

Theorem 4.12. *For any $f \in \mathcal{L}^1(X,\mu)$,*

$$\Big|\int_X f\,d\mu\Big| \leq \int_X |f|\,d\mu.$$

Proof. Let $a \in \mathbb{C}$ be such that $|a| = 1$ and

$$a \int_X f \, d\mu = \Big| \int_X f \, d\mu \Big|.$$

Then, writing $af = u + iv$, we have

$$\begin{aligned}
\Big| \int_X f \, d\mu \Big| &= a \int_X f \, d\mu \\
&= \int_X a \, f \, d\mu \\
&= \int_X u \, d\mu + \underbrace{i \int_X v \, d\mu}_{\text{=0, since lhs is real}} \\
&\le \int_X |u| \, d\mu \\
&\le \int_X |af| \, d\mu = \int_X |a| \, |f| \, d\mu \\
&= \int_X |f| \, d\mu
\end{aligned}$$

and the proof is complete. ∎

Theorem 4.13 (Lebesgue's Dominated Convergence Theorem). *Let (f_n) be a sequence of complex-valued measurable functions on X such that*

(i) *$f_n(x) \to f(x)$, as $n \to \infty$, for every $x \in X$ (pointwise convergence);*

(ii) *there is some $g \in \mathcal{L}^1(X, \mu)$ such that $|f_n(x)| \le g(x)$ for all $n \in \mathbb{N}$ and each $x \in X$ (the sequence (f_n) is dominated by g).*

Then $f \in \mathcal{L}^1(X, \mu)$ and $\int_X f_n \, d\mu \to \int_X f \, d\mu$ as $n \to \infty$. Furthermore, $\int_X |f_n - f| \, d\mu \to 0$, as $n \to \infty$.

Proof. Given $\varepsilon > 0$, set

$$E_n = \{ \, x \in X : |f_k(x) - f(x)| \le \varepsilon g(x) \text{ for all } k \ge n \, \}.$$

Then every $x \in X$ belongs to some E_n, that is, $\bigcup_n E_n = X$.
(Let $x \in X$. If $g(x) = 0$, then the hypotheses imply that $f_k(x) = 0$ for all k and so also $f(x) = 0$. Hence x belongs to every E_n. On the other hand, if $g(x) \ne 0$, then $f_k(x) \to f(x)$ implies that there is n_0 such that $|f_k(x) - f(x)| \le \varepsilon g(x)$ for all $k \ge n_0$, that is, $x \in E_{n_0}$.)

Evidently, $E_n \subseteq E_{n+1}$ so that (E_n) is an increasing sequence with union equal to X.

Next, $f = \lim_n f_n$ and so f is measurable. Furthermore, $|f| \le g$ and $g \in \mathcal{L}^1$ so that $f \in \mathcal{L}^1$. Now,

$$\begin{aligned}\Big|\int_X (f_n - f)\, d\mu\Big| &\le \int_X |f_n - f| \; d\mu \\ &= \int_{E_n} |f_n - f| \; d\mu + \int_{X \setminus E_n} |f_n - f| \; d\mu \\ &\le \int_{E_n} \varepsilon\, g \; d\mu + \int_{X \setminus E_n} 2g \; d\mu\end{aligned}$$

since $|f_n - f| \le \varepsilon g$ on E_n and $|f_n - f| \le 2g$ on X

$$\le \varepsilon \int_X g \; d\mu + \int_{X \setminus E_n} 2g \; d\mu.$$

But $E_n \uparrow X$ and so (using the Monotone Convergence Theorem applied to the sequence $((\mathbb{1}_{E_n} 2g))$, we see that $\int_{E_n} 2g\, d\mu \uparrow \int_X 2g\, d\mu$, as $n \to \infty$. It follows that $\int_{X \setminus E_n} 2g\, d\mu = \int_X (1 - \mathbb{1}_{E_n}) 2g\, d\mu \downarrow 0$. Therefore

$$\begin{aligned}\Big|\int_X f_n\, d\mu - \int_X f\, d\mu\Big| &= \Big|\int_X (f_n - f)\, d\mu\Big| \\ &\le \int_X |f_n - f| \; d\mu \\ &\to 0\end{aligned}$$

as $n \to \infty$ and the proof is complete. ■

Remark 4.14. Lebesgue's Dominated Convergence Theorem is one of the most important in the theory. It means that limits can be interchanged with integration (under the required conditions — which are fairly general).

Remark 4.15. Suppose that λ is Lebesgue measure on an interval $[a, b]$; thus, λ is the completion of the Lebesgue-Stieltjes measure on the Borel sets in $[a, b]$ determined by the function $F(x) = x$. Thus, $\lambda([a, t]) = t - a$ for any $a \le t \le b$. Then one can show the following.

Theorem 4.16. *Let f be a bounded real-valued function on $[a, b]$.*

(a) *The function f is Riemann-integrable on $[a, b]$ if and only if there is some set $A \subseteq [a, b]$ with $\lambda(A) = 0$ such that f is continuous at each point of $[a, b] \setminus A$.*

(b) *If f is Riemann-integrable on $[a, b]$, then f is integrable on $[a, b]$ with respect to λ and the two integrals are equal.*

Proof. This really just follows from the definitions. The Riemann integral is defined via upper and lower sums corresponding to partitions of $[a, b]$ — this can be rewritten in terms of simple functions. We omit the details here. ■

Theorem 4.17. *Let f be a bounded measurable function. Then $f \in \mathcal{L}^1(X, \mu)$. (Note: μ is a finite measure on X).*

Proof. Let $f = f_+ - f_-$, where $f_\pm$ are the positive and negative parts of f, respectively. Then there is some M such that $f_\pm(x) \le M$ for all $x \in X$. Hence, for any simple function s with $0 \le s(x) \le f_+(x)$, we have the inequality $\int_X s \, d\mu \le M \, \mu(X)$ and so it follows that (the supremum over all such simple functions is bounded by $M\mu(X)$ and therefore) f_+ is integrable. Similarly, f_- is integrable and so is f, i.e., $f \in \mathcal{L}^1(X, \mu)$. ∎

Remark 4.18. This result is not true if μ is not a finite measure. No non-zero constant is integrable with respect to Lebesgue measure on the real-line $\mathbb{R}$, but every constant is bounded.

Theorem 4.19 (Schwarz' Inequality). *Suppose that $|f|^2 \in \mathcal{L}^1(X, \mu)$ and $|g|^2 \in \mathcal{L}^1(X, \mu)$. Then $fg \in \mathcal{L}^1(X, \mu)$ and*

$$\Big| \int_X fg \, d\mu \Big| \le \int_X |fg| \, d\mu \le \Big(\int_X |f|^2 \, d\mu \Big)^{1/2} \Big(\int_X |g|^2 \, d\mu \Big)^{1/2}.$$

Proof. Set

$$A_n = \{ \, x \in X : |f(x)| \le n \, \} \quad \text{and} \quad B_n = \{ \, x \in X : |g(x)| \le n \, \}$$

and let $f_n = f \mathbb{1}_{A_n \cap B_n}$ and $g_n = g \mathbb{1}_{A_n \cap B_n}$. Then $|f_n| \le |f|$ and $|g_n| \le |g|$. However, f_n and g_n are bounded (by n) and therefore $f_n g_n$, f_n^2 and g_n^2 are bounded and are integrable, as is any linear combination of these functions. But then, for any $t \in \mathbb{R}$,

$$0 \le \int_X (t\,|f_n| + |g_n|)^2 \, d\mu = t^2 \int_X |f_n|^2 \, d\mu + 2t \int_X |f_n g_n| \, d\mu + \int_X |g_n|^2 \, d\mu$$

which means that

$$4 \Big(\int_X |f_n g_n| \, d\mu \Big)^2 \le 4 \int_X |f_n|^2 \, d\mu \int_X |g_n|^2 \, d\mu \qquad (*)$$

(the real quadratic $at^2 + bt + c \ge 0$ if and only if $b^2 \le 4ac$). But $|f_n| \uparrow |f|$ and $|g_n| \uparrow |g|$ and so $|f_n g_n| \uparrow |fg|$, as $n \to \infty$.

By Lebesgue's Dominated Convergence Theorem, the right hand side of inequality $(*)$ converges to $4 \int_X |f|^2 \, d\mu \int_X |g|^2 \, d\mu$. But then, looking at the left hand side, Lebesgue's Monotone Convergence Theorem shows that fg is integrable and the left hand side of inequality $(*)$ converges to $4\big(\int_X |fg| \, d\mu\big)^2$ from which the result follows. ∎

Chapter 5

Expectation in a probability space.

Let $(\Omega, \mathcal{S}, \mathbb{P})$ be a probability space and let $f : \Omega \to \mathbb{R}$ be a random variable. Denote by $\mathcal{L}^2(\Omega, \mathbb{P})$ the set $\{ f : |f|^2 \in \mathcal{L}^1(\Omega, \mathbb{P}) \}$ — the set of "square-integrable" functions on $(\Omega, \mathbb{P})$.

Definition 5.1. We say that the random variable f has finite expectation if $f \in \mathcal{L}^1(\Omega, \mathbb{P})$ and we set

$$\mathbb{E} f = \int_\Omega f \, d\mathbb{P}, \quad \text{the expectation of } f.$$

If $f \in \mathcal{L}^2(\Omega, \mathbb{P})$, the variance of f is

$$\operatorname{var} f = \int_\Omega (f - \mathbb{E} f)^2 \, d\mathbb{P} = \mathbb{E}(f - \mathbb{E} f)^2.$$

Remark 5.2. The constant random variable 1 is square-integrable and so, by Schwarz' inequality, if $f \in \mathcal{L}^2(\Omega, \mathbb{P})$, then $|f| \, 1 \in \mathcal{L}^1(\Omega, \mathbb{P})$, i.e., f is integrable. Therefore, $(f - \mathbb{E} f)^2 = f^2 - 2(\mathbb{E} f) f + (\mathbb{E} f)^2 \in \mathcal{L}^1(\Omega, \mathbb{P})$. In other words, the condition $f \in \mathcal{L}^2(\Omega, \mathbb{P})$ ensures that $\operatorname{var} f$ is well-defined.

We have seen that F_f, the distribution function of f, defines a measure on $\mathcal{B}(\mathbb{R})$ via the assignment $\mu((-\infty, a]) = F_f(a) = \operatorname{Prob}(f \le a)$. Indeed, $\mu(A) = \operatorname{Prob}(f \in A) = \mathbb{P}(f^{-1}(A))$ for any $A \in \mathcal{B}(\mathbb{R})$. The measure μ is a probability measure on $\mathbb{R}$ and so we can consider integrals, $\int_\mathbb{R} h \, d\mu$, over $\mathbb{R}$ with respect to μ. This is usually written as $\int_\mathbb{R} h \, dF_f$.

Theorem 5.3. *Suppose that f has finite expectation. Then*

$$\mathbb{E} f = \int_\mathbb{R} x \, dF_f.$$

Proof. To say that f has finite expectation is to say that f is integrable, i.e., $f \in \mathcal{L}^1(\Omega, \mathbb{P})$. We shall show that x is integrable with respect to μ on $\mathbb{R}$, that is, that $x \in \mathcal{L}^1(\mathbb{R}, \mu)$ (where μ is the measure on $\mathbb{R}$ given by F_f). Let s be any simple function on $\mathbb{R}$ with $0 \le s(x) \le |x|$; say,

$$s(x) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(x),$$

where $\{ A_1, \dots, A_n \}$ is a partition of $\mathbb{R}$. Then $s(x) \le |x|$ implies that if $x \in A_i$, then $s(x) = \alpha_i \le |x|$, that is, $\alpha_i \le |x|$ for all $x \in A_i$.

Let $B_i = \{ \omega : |f(\omega)| \in A_i \}$ and set

$$t(\omega) = \sum_{i=1}^n \alpha_i \mathbb{1}_{B_i}(\omega).$$

Note that $\{ B_1, \dots, B_n \}$ is a partition of Ω. Then t is a (measurable) simple function on Ω. Furthermore, if $\omega \in B_i$, say, then

$$t(\omega) = \alpha_i \le |f(\omega)|,$$

since $|f(\omega)| \in A_i$. Hence $0 \le t \le |f|$ and so $\int_\Omega t \, d\mathbb{P} \le \int_\Omega |f| \, d\mathbb{P}$. But

$$\begin{aligned}
\int_{\mathbb{R}} s \, d\mu &= \sum_{i=1}^n \alpha_i \, \mu(A_i), \quad \text{by definition of the integral,} \\
&= \sum_{i=1}^n \alpha_i \, \mathbb{P}(B_i), \quad \text{by definition of the measure } \mu, \\
&= \int_\Omega t \, d\mathbb{P} \\
&\le \int_\Omega |f| \, d\mathbb{P}.
\end{aligned}$$

Hence the set $\{ \int_{\mathbb{R}} s \, d\mu : 0 \le s(x) \le |x|, \quad s \text{ simple} \}$ is bounded and therefore $|x|$ is integrable with respect to μ.

We wish to show that $\mathbb{E} f = \int_{\mathbb{R}} x \, d\mu$. For $k = 1, 2, \dots,$ let I_k denote the interval $I_k = ((k-1)/n, k/n]$ and let J_k denote the interval $J_k = (-k/n, -(k-1)/n]$. Set

$$s_n(x) = \sum_{k=1}^{n^2} \big(\tfrac{k-1}{n}\big) \mathbb{1}_{I_k}(x) + n \mathbb{1}_{(n,\infty)}(x) - \sum_{m=1}^{n^2} \big(\tfrac{m-1}{n}\big) \mathbb{1}_{J_m}(x) - n \mathbb{1}_{(-\infty,n]}(x).$$

We see that $|s_n(x)| \le |x|$ for all n. Let

$$\begin{aligned}
A_k &= \{ \omega : f(\omega) \in I_k \} = f^{-1}(A_k), \\
B_k &= \{ \omega : f(\omega) \in J_k \} = f^{-1}(B_k), \\
A'_n &= \{ \omega : f(\omega) \ge n \} = f^{-1}((n,\infty)), \\
B'_n &= \{ \omega : f(\omega) \le -n \} = f^{-1}((\infty,-n]).
\end{aligned}$$

For any given $x \in \mathbb{R}$, $-n < x < n$ for all sufficiently large n. Hence, for large n, $x \notin (\infty, n] \cup [n, \infty)$ and therefore $|s_n(x) - x| \le 1/n$. In other words, $s_n(x) \to x$, as $n \to \infty$, for each given $x \in \mathbb{R}$.

Set

$$t_n(\omega) = \sum_{k=1}^{n^2} \big(\tfrac{k-1}{n}\big)\, \mathbb{1}_{A_k}(\omega) + n\,\mathbb{1}_{A'_n}(\omega) - \sum_{m=1}^{n^2} \big(\tfrac{m-1}{n}\big)\, \mathbb{1}_{B_m}(\omega) - n\,\mathbb{1}_{B'_n}(\omega).$$

Then, for given $\omega \in \Omega$, $\omega \notin A'_n \cup B'_n$ for all sufficiently large n so that $|t_n(\omega) - f(\omega)| \le 1/n$. In other words, $t_n(\omega) \to f(\omega)$ on Ω, as $n \to \infty$. Furthermore, by construction, $|t_n(\omega)| \le |f(\omega)|$ for all $\omega \in \Omega$. Therefore, by Lebesgue's Dominated Convergence Theorem,

$$\int_\Omega t_n \, d\mathbb{P} \to \int_\Omega f \, d\mathbb{P}.$$

But, again by the Dominated Convergence Theorem,

$$\int_\mathbb{R} s_n \, d\mu \to \int_\mathbb{R} x \, d\mu$$

and, since $\int_\mathbb{R} s_n \, d\mu = \int_\Omega t_n \, d\mathbb{P}$, we conclude that $\int_\mathbb{R} x \, d\mu = \int_\Omega f \, d\mathbb{P}$. ∎

We can generalize this result somewhat.

Theorem 5.4. *Let $g : \mathbb{R} \to \mathbb{R}$ be Borel measurable and let X be a random variable. Then*

$$\mathbb{E}(g(X)) = \int_\mathbb{R} g(x) \, dF_X$$

in that if either side exists then so does the other and they are equal.

Proof. Suppose first that g has the form $g(x) = \alpha \mathbb{1}_A(x)$ where $\alpha \ge 0$ and $A \in \mathcal{B}(\mathbb{R})$. Let $B = \{\, \omega : X(\omega) \in A \,\}$. Then $B \in \mathcal{S}$. We have

$$g(X)(\omega) = g(X(\omega)) = \alpha \mathbb{1}_A(X(\omega)) = \alpha \mathbb{1}_B(\omega).$$

Hence $\int_\Omega g(X) \, d\mathbb{P} = \alpha \mathbb{P}(B)$. But

$$\int_\mathbb{R} g(x) \, dF_X = \alpha\, \mu(A) = \alpha\, \mathbb{P}(B)$$

by definition of the measure μ determined by F_X. So

$$\mathbb{E}(g(X)) = \int_\mathbb{R} g(x) \, dF_X$$

in this case. Next, suppose that g is a non-negative simple function on $\mathbb{R}$, say, $g(x) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(x)$, with $\alpha_i \ge 0$ for all $1 \le i \le n$. Then

$$\begin{aligned} \mathbb{E}(g(X)) &= \mathbb{E}\big(\textstyle\sum_i \alpha_i \mathbb{1}_{A_i}(X)\big) \\ &= \sum_i \alpha_i \mathbb{E}(\mathbb{1}_{A_i}(X)) \end{aligned}$$

$$
\begin{aligned}
&= \sum_i \int_{\mathbb{R}} \alpha_i \mathbb{1}_{A_i} \, dF_X \,, \quad \text{from the result above,} \\
&= \int_{\mathbb{R}} g(x) \, dF_X
\end{aligned}
$$

and we see that the theorem is true for such g.

Now let $g \geq 0$ be any non-negative Borel function and let (g_n) be an increasing sequence of simple functions on $\mathbb{R}$ with $0 \leq g_n(x) \leq g(x)$ and $g_n(x) \to g(x)$ for each $x \in \mathbb{R}$. By the above, $\int_\Omega g_n(X) \, d\mathbb{P} = \int_{\mathbb{R}} g_n(x) \, dF_X$. But then, by the Monotone Convergence Theorem, we deduce that

$$\int_\Omega g(X) \, d\mathbb{P} = \int_{\mathbb{R}} g(x) \, dF_X$$

— if either side exists, so does the other, and they are equal.

Finally, for arbitrary Borel $g : \mathbb{R} \to \mathbb{R}$, we write $g = g_+ - g_-$ where $g_\pm \geq 0$. Then, as above,

$$\int_\Omega g_\pm(X) \, d\mathbb{P} = \int_{\mathbb{R}} g_\pm \, dF_X .$$

If both $\int_\Omega g_\pm(X) \, d\mathbb{P}$ exist, then so does their sum

$$\int_\Omega g_+(X) \, d\mathbb{P} + \int_\Omega g_-(X) \, d\mathbb{P} = \int_\Omega |g(X)| \, d\mathbb{P}.$$

Also, both $\int_{\mathbb{R}} g_\pm \, dF_X$ exist and so $|g|$ is integrable with respect to μ and we have

$$\int_\Omega (g_+(X) - g_-(X)) \, d\mathbb{P} = \int_{\mathbb{R}} (g_+ - g_-) \, dF_X ,$$

that is,

$$\mathbb{E}(g(X)) = \int_\Omega g(X) \, d\mathbb{P} = \int_{\mathbb{R}} g \, dF_X$$

and the proof is complete. ■

Corollary 5.5. *For a random variable X,*

$$\operatorname{var} X = \int_{\mathbb{R}} (x - \mathbb{E} X)^2 \, dF_X .$$

Proof. Let $g(x) = (x - \mathbb{E} X)^2$ and apply the theorem. ■

Theorem 5.6. *Suppose that the random variable has distribution function F_X given by*

$$F_X(a) = \int_{(-\infty, a]} \varphi(x) \, dx$$

for some non-negative Borel function φ (with $\int_{\mathbb{R}} \varphi(x)\,dx = 1$), where dx denotes Lebesgue measure on $\mathbb{R}$. Then for any Borel function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x)\,\varphi(x)\,dx.$$

Proof. If μ denotes the Lebesgue-Stieltjes measure on $\mathbb{R}$ given by F_X, then the hypothesis on F_X means that $\mu(A) = \int_A \varphi(x)\,dx$ for any Borel set A in $\mathbb{R}$. Now, suppose that $g(x) = \alpha \mathbb{1}_A(x)$ for $\alpha \geq 0$ and $A \in \mathcal{B}(\mathbb{R})$. Then we have

$$\begin{aligned}\mathbb{E}(g(X)) &= \int_{\mathbb{R}} \alpha \mathbb{1}_A(x)\,dF_X \\ &= \alpha\mu(A) \\ &= \alpha \int_A \varphi(x)\,dx \\ &= \int_{\mathbb{R}} \alpha \mathbb{1}_A(x)\varphi(x)\,dx.\end{aligned}$$

By linearity, it follows that

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x)\,\varphi(x)\,dx$$

for any non-negative simple (Borel) function g on $\mathbb{R}$. But if (g_n) is an increasing sequence of such functions, converging pointwise to $g \geq 0$, then $(g_n\varphi)$ is an increasing sequence of non-negative functions converging pointwise to $g\varphi$. By the Monotone Convergence Theorem, (applied in both $(\Omega, \mathcal{S}, \mathbb{P})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}), dx)$), we have

$$\int_\Omega g_n(X)\,d\mathbb{P} \to \int_\Omega g(X)\,d\mathbb{P}$$

and

$$\int_{\mathbb{R}} g_n(x)\,\varphi(x)\,dx \to \int_{\mathbb{R}} g(x)\,\varphi(x)\,dx$$

as $n \to \infty$. However, for each n the left hand sides are equal and so therefore are their limits, i.e.,

$$\mathbb{E}(g(X)) = \int_\Omega g(X)\,d\mathbb{P} = \int_{\mathbb{R}} g(x)\,\varphi(x)\,dx$$

holds for any (Borel) function $g \geq 0$. The general result now follows (by linearity) by writing an arbitrary (Borel) function g as $g = g_+ - g_-$. ∎

Remark 5.7. If $g(x)\varphi(x)$ is Riemann-integrable, then $\int_{\mathbb{R}} g(x)\varphi(x)dx$ has the same value irrespective of whether it is considered as a Lebesgue or as a Riemann integral and so we recover the standard formula from elementary probability theory.

Example 5.8. The random variable X has $N(\mu, \sigma^2)$ distribution (i.e., normal with mean μ and standard deviation σ) if F_X is given by

$$F_X(a) = \int_{-\infty}^{a} e^{-(x-\mu)^2/2\sigma^2} \frac{dx}{\sigma\sqrt{2\pi}}\,.$$

Then

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)\, e^{-(x-\mu)^2/2\sigma^2} \frac{dx}{\sigma\sqrt{2\pi}}$$

— where this is a Riemann integral provided that g is sufficiently well-behaved (for example, if g is continuous and polynomially bounded).

Definition 5.9. A random variable X is said to be absolutely continuous if its distribution function F_X has the form

$$F_X(x) = \int_{-\infty}^{x} \varphi(t)\, dt$$

for some non-negative function φ on $\mathbb{R}$ with $\int_{\mathbb{R}} \varphi(t)\, dt = 1$. If there is no risk of confusion, one often simply just says that X is a 'continuous' random variable rather than 'absolutely continuous'.

If X is (absolutely) continuous, then $F_X : \mathbb{R} \to [0,1]$ is a continuous function. To see this, fix $x \in \mathbb{R}$ and suppose that $a_n < x$ for all n and that $a_n \to x$. We have

$$F(x) - F(a_n) = \int_{(a_n,x]} \varphi(t)\, dt = \int_{\mathbb{R}} \mathbb{1}_{(a_n,x]}(t)\, \varphi(t)\, dt\,.$$

Let $g_n(t) = \mathbb{1}_{(a_n,x]}(t)\, \varphi(t)$. Then $0 \le g_n(t) \le \varphi(t)$ and $g_n(t) \to \mathbb{1}_{\{x\}}(t)\, \varphi(x)$, so by the Dominated Convergence Theorem,

$$\int_{\mathbb{R}} g_n(t)\, dt \to \int_{\mathbb{R}} \mathbb{1}_{\{x\}}(t)\, \varphi(x)\, dt = 0.$$

This shows that F_X is continuous at x from the left. Since every distribution function is continuous from the right, we conclude that F_X is, indeed, continuous on $\mathbb{R}$, as claimed.

The converse is *false*. For example, if F is the Cantor distribution, then we have seen that $F : \mathbb{R} \to [0,1]$ is continuous but $F(x)$ *cannot* be written as $\int_{-\infty}^{x} \varphi(t)\, dt$ for any φ.

Chapter 6

Characteristic functions.

Definition 6.1. Let X be a random variable on a probability space $(\Omega, \mathcal{S}, \mathbb{P})$. The characteristic function of X is the function $\varphi_X : \mathbb{R} \to \mathbb{C}$ given by

$$\varphi_X(t) = \int_\Omega e^{itX}\, d\mathbb{P} = \mathbb{E}(e^{itX}) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX))$$

for $t \in \mathbb{R}$.

Note that both $\cos(tX)$ and $\sin(tX)$ are bounded (Borel measurable) functions on Ω and so are integrable for all $t \in \mathbb{R}$. The moment generating function, $M_X(t)$ of X has a similar definition, but without the i, namely, $M_X(t) = \int_\Omega e^{tX}\, d\mathbb{P} = \mathbb{E}(e^{tX})$. Unfortunately, it may happen that this is not finite for any non-zero t. (For example, this happens if X has the Cauchy distribution, $F_X(x) = \int_{-\infty}^x ds/\pi(1+s^2)$. In this case, $M_X(t) = \int_\mathbb{R} \big(e^{st}/\pi(1+s^2) \big)\, ds$.)

Theorem 6.2. *The characteristic function φ_X satisfies the following:*

(i) $\left|\varphi_X(t)\right| \le \varphi_X(0) = 1$.

(ii) φ_X *is uniformly continuous on* $\mathbb{R}$.

(iii) $\overline{\varphi_X(t)} = \varphi_X(-t)$.

Proof. (i) Clearly, $\varphi_X(0) = 1$. Furthermore,

$$\left|\varphi_X(t)\right| = \Big| \int_\Omega e^{itX}\, d\mathbb{P} \Big| \le \int_\Omega \left|e^{itX}\right|\, d\mathbb{P} = \int_\Omega d\mathbb{P} = 1.$$

(ii) For any $s, t \in R$,

$$\begin{aligned}\left|\varphi_X(t) - \varphi_X(s)\right| &\le \int_\Omega \left|e^{itX} - e^{isX}\right|\, d\mathbb{P} = \int_\Omega \left|e^{isX}\right| \left|e^{i(t-s)X} - 1\right| d\mathbb{P} \\ &= \int_\Omega \left|e^{i(t-s)X} - 1\right| d\mathbb{P}\,.\end{aligned}$$

But for each $\omega \in \Omega$, $\left|e^{i\delta_n X(\omega)} - 1\right| \to 0$ as $n \to \infty$ for any sequence (δ_n) with $\delta_n \to 0$. By the Dominated Convergence Theorem, we deduce that φ_X is uniformly continuous on $\mathbb{R}$.

(iii) This is clear form the definition. ∎

Theorem 6.3. *Suppose that* $\mathbb{E}(|X|^n)$ *exists for some* $n \geq 1$ *(i.e.,* $X^n \in \mathcal{L}^1$*). Then, for all* $1 \leq r \leq n$*, the* r^{th}*-derivative* $\varphi_X^{(r)}(t)$ *exists and*

$$\varphi_X^{(r)}(t) = \int_\Omega (iX)^r e^{itX}\, d\mathbb{P} = \mathbb{E}\big((iX)^r e^{itX}\big)$$

and

$$\mathbb{E}(X^r) = \frac{\varphi_X^{(r)}(0)}{i^r}.$$

Proof. If $X^n \in \mathcal{L}^1$, then it follows that $X^r \in \mathcal{L}^1$ for all $r \leq n$. Now

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \mathbb{E}\Big(e^{itX}\Big(\frac{e^{ihX} - 1}{h}\Big)\Big),$$

and $\big|(e^{ihX(\omega)} - 1)/h\big| \leq X(\omega)$ for each $\omega \in \Omega$. Moreover, if $h_n \to 0$ as $n \to \infty$, then $(e^{ih_n X(\omega)} - 1)/h_n \to X(\omega)$ as $n \to \infty$ so, by the Dominated Convergence Theorem, we deduce that

$$\mathbb{E}\Big(e^{itX}\Big(\frac{e^{ih_n X} - 1}{h_n}\Big)\Big) \to \mathbb{E}(iXe^{itX})$$

as $n \to \infty$. Hence $\varphi_X(t)$ is differentiable with $\varphi_X^{(1)}(t) = \mathbb{E}(iXe^{itX})$.

The general result follows by induction. Finally, putting $t = 0$ gives the formula $\mathbb{E}(X^r) = \varphi_X^{(r)}(0)/i^r$. ∎

Remark 6.4. We can express these formulae in terms of F_X as follows.

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \iint_\Omega e^{itX}\, d\mathbb{P} = \int_\mathbb{R} e^{itx}\, dF_X,$$

and

$$\mathbb{E}(X^n) = \int_\mathbb{R} x^n\, dF_X = \frac{\varphi_X^{(n)}(0)}{i^n}.$$

We recall the following approximation theorem (Weierstrass).

Proposition 6.5. *Let* g *be a periodic continuous function on* $\mathbb{R}$ *with period* 2π*. Then* g *can be uniformly approximated by trigonometric polynomials: i.e., given* $\varepsilon > 0$*, there is some* $m \in \mathbb{N}$ *and* $\alpha_0, \alpha_1, \ldots, \alpha_m$ *such that*

$$\Big|\, g(x) - \sum_{j=0}^{m} \alpha_j e^{ijx} \,\Big| < \varepsilon$$

for all $x \in \mathbb{R}$.
N.B. If h *is continuous and periodic with period* $2n$*, then defining* g *by*

$g(y) = h(ny/\pi)$, we see that g is continuous and periodic with period 2π. Hence, as above, $\left|g(y) - \sum_{j=0}^{m} \alpha_j e^{ijy}\right| < \varepsilon$ for all y. Putting $y = \pi x/n$ we get

$$\Big|\, h(x) - \sum_{j=0}^{m} \alpha_j e^{ixj\pi/n} \,\Big| < \varepsilon$$

for all $x \in \mathbb{R}$.

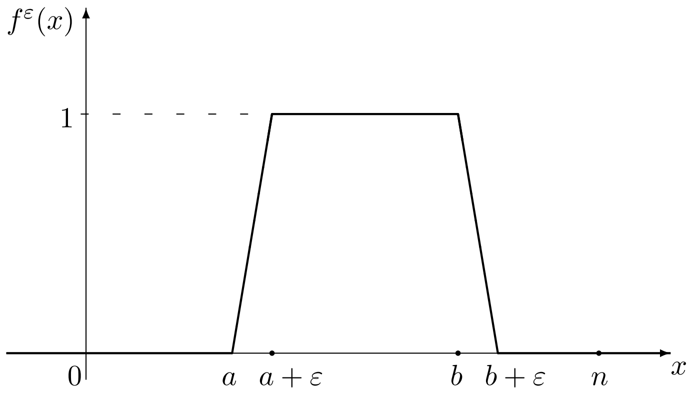
Theorem 6.6. *Let X and Y be random variables with distribution functions F and G, respectively. Then $\varphi_X(t) = \varphi_Y(t)$ for all t if and only if $F(x) = G(x)$ for all $x \in \mathbb{R}$.*

Proof. If $F = G$, then

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx}\, dF = \int_{\mathbb{R}} e^{itx}\, dG = \varphi_Y(t)$$

for all $t \in \mathbb{R}$.

Conversely, suppose that $\varphi_X(t) = \varphi_Y(t)$ for all $t \in \mathbb{R}$. Let $a < b$ be fixed and let $\varepsilon > 0$ be given. Let n be so large that $[a, b+\varepsilon] \subset [-n, n]$ and let f^ε be the (piece-wise linear) function on $\mathbb{R}$ as shown in the diagram.



Now, $f^\varepsilon(-n) = f^\varepsilon(n)$ $(= 0)$ so f^ε can be approximated uniformly on $[-n, n]$ by trigonometric polynomials. Thus, with $\varepsilon' = 1/n$, there is $\alpha_0, \alpha_1, \dots, \alpha_m$ such that

$$|f^\varepsilon(x) - p_n(x)| < \varepsilon' = 1/n$$

for all $x \in [-n, n]$, where $p_n(x)$ denotes the polynomial $\sum_{j=1}^{m} \alpha_j e^{ijx\pi/n}$. Now, $|f^\varepsilon| \le 1$ and so $|p_n(x)| \le 2$ for any $x \in [-n, n]$, and so, by periodicity, $|p_n(x)| \le 2$ for all $x \in \mathbb{R}$. Furthermore,

$$\int_{\mathbb{R}} p_n(x)\, dF = \int_{\mathbb{R}} \textstyle\sum_{j=1}^{m} \alpha_j\, e^{ijx\pi/n}\, dF$$

$$
\begin{aligned}
&= \sum_{j=1}^{m} \alpha_j \, \varphi_X(j\pi/n) \\
&= \sum_{j=1}^{m} \alpha_j \, \varphi_Y(j\pi/n)\,, \quad \text{since } \varphi_X = \varphi_Y, \\
&= \int_{\mathbb{R}} p_n(x)\, dG.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\Big| \int_{\mathbb{R}} f^\varepsilon \, dF - \int_{\mathbb{R}} f^\varepsilon \, dG \Big| &= \Big| \int_{\mathbb{R}} \mathbb{1}_{[-n,n]} f^\varepsilon \, dF - \int_{\mathbb{R}} \mathbb{1}_{[-n,n]} f^\varepsilon \, dG \Big| \\
&= \Big| \int_{\mathbb{R}} \mathbb{1}_{[-n,n]} (f^\varepsilon - p_n)\, dF - \int_{\mathbb{R}} \mathbb{1}_{[-n,n]} (f^\varepsilon - p_n)\, dG \\
&\qquad + \int_{\mathbb{R}} \mathbb{1}_{[-n,n]} p_n \, dF - \int_{\mathbb{R}} \mathbb{1}_{[-n,n]} p_n \, dG \Big| \\
&\le \int_{\mathbb{R}} \mathbb{1}_{[-n,n]} \, |f^\varepsilon - p_n| \; dF + \int_{\mathbb{R}} \mathbb{1}_{[-n,n]} \, |f^\varepsilon - p_n| \; dG \\
&\qquad + \Big| \int_{\mathbb{R}} \mathbb{1}_{[-n,n]} p_n \, dF - \int_{\mathbb{R}} \mathbb{1}_{[-n,n]} p_n \, dG \Big| \\
&\le \frac{2}{n} + \Big| \int_{\mathbb{R}} p_n \, dF - \int_{\mathbb{R}} p_n \, dG \Big| \\
&\qquad + \int_{\mathbb{R}} (1 - \mathbb{1}_{[-n,n]})\, |p_n| \; dF + \int_{\mathbb{R}} (1 - \mathbb{1}_{[-n,n]})\, |p_n| \; dG \\
&\le \frac{2}{n} + 0 + 2\mu_F(\mathbb{R} \setminus [-n,n]) + 2\mu_G(\mathbb{R} \setminus [-n,n]) \\
&\to 0 \quad \text{as } n \to \infty.
\end{aligned}
$$

It follows that $\int_{\mathbb{R}} f^\varepsilon \, dF = \int_{\mathbb{R}} f^\varepsilon \, dG$. But as $\varepsilon \downarrow 0$, $f^\varepsilon \to \mathbb{1}_{(a,b]}$ and so by the Dominated Convergence Theorem, applied to a sequence $\varepsilon_n \downarrow 0$, we deduce that

$$
\int_{\mathbb{R}} \mathbb{1}_{(a,b]} \, dF = \int_{\mathbb{R}} \mathbb{1}_{(a,b]} \, dG\,,
$$

that is, $F(b) - F(a) = G(b) - G(a)$. Letting $a \to -\infty$, we find $F(b) = G(b)$ for any $b \in \mathbb{R}$ and the proof is complete. ∎

Remark 6.7. It is possible for two characteristic functions to agree on some interval in $\mathbb{R}$, but not on the whole of $\mathbb{R}$.

There is an inversion theorem, which we will not prove quite completely.

Theorem 6.8. *Let X be a random variable with distribution function F and let $\overline{F} = \frac{1}{2}\{F(x) + F(x-)\}$. Then, for any $a \le b$,*

$$
\overline{F}(b) - \overline{F}(a) = \lim_{A \to \infty} \frac{1}{2\pi} \int_{-A}^{A} \frac{(e^{-iat} - e^{-ibt})}{it} \; \varphi_X(t)\, dt.
$$

Proof. We substitute $\varphi_X(t) = \int_{\mathbb{R}} e^{itx}\, dF_X$ and then change the order of integration (which can be justified).

$$\begin{aligned} I_A &\equiv \frac{1}{2\pi} \int_{-A}^{A} \Big(\frac{e^{-iat} - e^{-ibt}}{it} \Big) \int_{\mathbb{R}} e^{itx}\, dF_X \; dt \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} dF_X \int_{-A}^{A} \Big(\frac{e^{it(x-a)} - e^{it(x-b)}}{it} \Big)\, dt \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} dF_X \; 2 \int_{A(x-b)}^{A(x-a)} \frac{\sin y}{y}\, dy \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} dF_X \; 2\{ g(A(x-a)) - g(A(x-b)) \} \end{aligned}$$

where $g(s) = \int_0^s \sin y / y\, dy$. Now, g is an odd function and $g(s) \to \frac{1}{2}\pi$ as $s \to \infty$ and therefore g is bounded and I_A is given by

$$I_A = \frac{1}{\pi} \int_{\mathbb{R}} \{ g(A(x-a)) - g(A(x-b)) \}\, dF_X .$$

As $A \to \infty$, the integrand converges to

$$\begin{cases} 0 & x < a \\ \pi/2 & x = a \\ \pi & a < x < b \\ \pi/2 & x = b \\ 0 & x > b. \end{cases}$$

Hence, by the Dominated Convergence Theorem,

$$\begin{aligned} \lim_{A \to \infty} I_A &= \int_{\mathbb{R}} (\mathbb{1}_{\{a\}} + \mathbb{1}_{(a,b)} + \tfrac{1}{2} \mathbb{1}_{\{b\}})\, dF_X \\ &= \tfrac{1}{2}(F(a) - F(a-)) + (F(b-) - F(a)) + \tfrac{1}{2}(F(b) - F(b-)) \\ &= \tfrac{1}{2}(F(b) - F(b-)) - \tfrac{1}{2}(F(a) - F(a-)) \\ &= \overline{F}(b) - \overline{F}(a)\, , \end{aligned}$$

as required. ∎

Remark 6.9. Suppose that $\int_{-\infty}^{\infty} |\varphi_X(t)|\, dt$ exists. Then putting $a = 0$ in the formula (and interchanging orders of integration), we get

$$\begin{aligned} \overline{F}(b) - \overline{F}(0) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Big(\frac{1 - e^{-ibt}}{it} \Big)\, \varphi_X(t)\, dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(t)\, dt \int_0^b e^{-itx}\, dx \end{aligned}$$

$$= \frac{1}{2\pi} \int_0^b dx \int_{-\infty}^{\infty} \varphi_X(t)\, e^{-itx}\, dt\,.$$

From this it follows that X has probability density function

$$p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(t)\, e^{-itx}\, dt\,.$$

(Thus, the integrability of φ_X implies that X is absolutely continuous and has probability density function given by the Fourier transform of φ_X.)

Examples 6.10.

1. If X is $B(n,p)$, then $\varphi_X(t) = (q + pe^(it))^n$, where $q = 1 - p$.
2. If X is a Poisson random variable with mean μ, then the characteristic function of X is given by $\varphi_X(t) = \exp\big(\mu(e^{it} - 1)\big)$.
3. If X is a normal random variable, $N(\mu, \sigma^2)$, then $\varphi_X(t) = e^{it\mu}\, e^{-\sigma^2 t^2/2}$.

Chapter 7

Independence.

We recall that two events A and B are independent if

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

We wish to extend the idea of independence from events to random variables.

Definition 7.1. Random variables $X_1, \ldots, X_n$ are said to be independent if for any Borel sets $E_1, \ldots, E_n$ in $\mathbb{R}$, we have

$$\mathrm{Prob}(X_1 \in E_1, \ldots, X_n \in E_n) = \mathrm{Prob}(X_1 \in E_1) \ldots \mathrm{Prob}(X_n \in E_n)$$

i.e., if $\mathbb{P}(X_1^{-1}(E_1) \cap \cdots \cap X_n^{-1}(E_n)) = \prod_{j=1}^n \mathbb{P}(X_j^{-1}(E_j))$.

Hence, events A and B are independent if and only if the random variables $\mathbb{1}_A$ and $\mathbb{1}_B$ are independent.

Theorem 7.2. *The random variables $X_1, \ldots, X_n$ are independent if and only if*

$$\mathbb{P}(X_1 \le a_1, \ldots, X_n \le a_n) = \mathbb{P}(X_1 \le a_1) \ldots \mathbb{P}(X_n \le a_n)$$

for all $a_1, \ldots, a_n \in \mathbb{R}$.

Proof. If $X_1, \ldots, X_n$ are independent, then the result follows directly by setting $E_1 = (-\infty, a_1], \ldots, E_n = (-\infty, a_n]$ for any given $a_1, \ldots, a_n \in \mathbb{R}$.

Conversely, suppose that

$$\mathbb{P}(X_1 \le a_1, \ldots, X_n \le a_n) = \mathbb{P}(X_1 \le a_1) \ldots \mathbb{P}(X_n \le a_n)$$

for all $a_1, \ldots, a_n \in \mathbb{R}$. Fix $a_2, \ldots, a_n$ and for any $E \in \mathcal{B}(\mathbb{R})$, set

$$\mu(E) = \mathbb{P}(X_1 \in E, X_2 \le a_2, \ldots, X_n \le a_n)$$

and

$$\nu(E) = \mathbb{P}(X_1 \in E)\,\mathbb{P}(X_2 \le a_2) \ldots \mathbb{P}(X_n \le a_n).$$

Then we see that μ and ν are finite measures on $\mathcal{B}(\mathbb{R})$. Moreover, for any $a \in \mathbb{R}$,

$$\begin{aligned}\mu((-\infty,a]) &= \mathbb{P}(X_1 \le a, X_2 \le a_2, \dots, X_n \le a_n)\\ &= \mathbb{P}(X_1 \le a)\mathbb{P}(X_2 \le a_2)\dots\mathbb{P}(X_n \le a_n)\\ &= \nu((-\infty,a])\,,\end{aligned}$$

so that $\mu((-\infty,a]) = \nu((-\infty,a])$. Letting $a \to \infty$, we see that $\mu(\mathbb{R}) = \nu(\mathbb{R})$. Hence, for any α, β,

$$\begin{aligned}\mu((\alpha,\beta]) &= \mu((-\infty,\beta]) - \mu((-\infty,\alpha])\\ &= \nu((-\infty,\beta]) - \nu((-\infty,\alpha])\\ &= \nu((\alpha,\beta]).\end{aligned}$$

Also,

$$\begin{aligned}\mu((a,\infty)) &= \mu((\mathbb{R}) - \mu((-\infty,a])\\ &= \nu((\mathbb{R}) - \nu((-\infty,a])\\ &= \nu((a,\infty)).\end{aligned}$$

Let $\mathcal{A}$ be the algebra of sets in $\mathbb{R}$ generated by the intervals $(-\infty,a]$ with $a \in \mathbb{R}$. $\mathcal{A}$ can be described explicitly as follows. Let $I(a)$ denote the interval $(-\infty,a]$ and, for $a < b$, let $J(a,b)$ denote the interval $(a,b]$ ($= I(a)^c \cup I(b)$). Then $\mathcal{A}$ consists of those subsets of $\mathbb{R}$ which take one of the following forms. $\mathbb{R}$ or $\varnothing$ or $I(a_1) \cup J(a_2,b_2) \cup \dots \cup J(a_n,b_n)$ or $J(a_1,b_1) \cup \dots \cup J(a_m,b_m)$ or $J(a_1,b_1) \cup \dots \cup J(a_k,b_k) \cup (\alpha,\infty)$ or $I(a_1) \cup J(a_2,b_2) \cup \dots \cup J(a_r,b_r) \cup (\alpha,\infty)$ where the sets in each finite union may be taken disjoint. Evidently, $\mu(E) = \nu(E)$ for any $E \in \mathcal{A}$.

Let $\mathcal{M} = \{\, E \in \mathcal{B}(\mathbb{R}) : \mu(E) = \nu(E)\,\}$. If $A_1 \subseteq A_2 \subseteq \dots$ belong to $\mathcal{M}$, then

$$\mu\big(\bigcup_i A_i\big) = \lim_n \mu(A_n) = \lim_n \nu(A_n) = \nu\big(\bigcup_i A_i\big)$$

and therefore $\bigcup_i A_i \in \mathcal{M}$. Similarly, if $B_1 \supseteq B_2 \supseteq \dots$ belong to $\mathcal{M}$, then we see that $\bigcap_i B_i \in \mathcal{M}$. Thus, $\mathcal{M}$ is a monotone class. Since $\mathcal{A} \subseteq \mathcal{M}$, it follows that $\Sigma(\mathcal{A}) \subseteq \mathcal{M} \subseteq \mathcal{B}(\mathbb{R})$. But $\Sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$ and so $\mathcal{M} = \mathcal{B}(\mathbb{R})$.

Now fix E_1 in $\mathcal{B}(\mathbb{R})$ and start again.

For $E \in \mathcal{B}(\mathbb{R})$, set $\mu(E) = \mathbb{P}(X_1 \in E_1, X_2 \in E, X_3 \le a_3, \dots, X_n \le a_n)$ and set $\nu(E) = \mathbb{P}(X_1 \in E_1)\mathbb{P}(X_2 \in E)\mathbb{P}(X_3 \le a_3)\dots\mathbb{P}(X_n \le a_n)$. Then, as above (and using the above), one shows that $\mu = \nu$ on $\mathcal{A}$ and, once again, it follows that $\mu = \nu$ on $\mathcal{B}(\mathbb{R}) = \Sigma(\mathcal{A})$.

Repeating this (i.e., by induction) the result follows. ∎

Definition 7.3. For random variables $X_1, \dots, X_n$, their joint distribution function $F_{X_1,\dots,X_n} : \mathbb{R}^n \to [0,1]$ is the function

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) = \mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n).$$

One can show that $F_{X_1,\dots,X_n}$ satisfies properties similar to those satisfied by F_X, the distribution function of the single random variable X (such as monotonicity and right continuity). We note the following (consistency relation)

$$\lim_{x_j \to \infty} F_{X_1,\dots,X_n}(x_1,\dots,x_n) = F_{X_1,\dots,\widehat{X_j},\dots,X_n}(x_1,\dots,\widehat{x_j},\dots,x_n),$$

where $\widehat{y}$ means omit the term y.

Corollary 7.4. *The random variables $X_1,\dots,X_n$ are independent if and only if*

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) = F_{X_1}(x_1)\dots F_{X_n}(x_n)$$

for all $x_1,\dots,x_n \in \mathbb{R}$.

Proof. This follows immediately from Theorem 7.2 and the definitions. ∎

Remark 7.5. Just as for a single random variable, one can show that given a function $F : \mathbb{R}^n \to [0,1]$ satisfying certain conditions (those which a joint distribution function must satisfy) then there are random variables whose joint distribution function *is* the given function F. In fact, one shows that F defines a probability measure $\mathbb{P}_F$ (the Lebesgue-Stieltjes measure) on the Borel sets in $\mathbb{R}^n$. Then one takes $\Omega = \mathbb{R}^n$, $\mathcal{S} = \mathcal{B}(\mathbb{R}^n)$ and puts $X_i(\omega) = x_i$ where $\omega = (x_1,\dots,x_n) \in \Omega = \mathbb{R}^n$. By construction, F is the joint distribution of the $X_1,\dots,X_n$.

This shows that there do exist random variables with arbitrary preassigned joint distribution. In particular, suppose that $F_1,\dots,F_n$ are n given distribution functions. Then $F(x_1,\dots,x_n) = F_1(x_1)\dots F_n(x_n)$ *is* a joint distribution function for random variables $X_1,\dots,X_n$ and these are independent. So independent random variables with arbitrary preassigned individual distributions always exist.

Theorem 7.6. *Suppose that $X_1,\dots,X_n$ are independent random variables and that $g_1,\dots,g_n$ are Borel functions on $\mathbb{R}$. Then the random variables $Y_1 = g_1(X_1),\dots,Y_n = g_n(X_n)$ are independent.*

Proof. Let $E_1,\dots,E_n$ be Borel sets in $\mathbb{R}$. Then

$$\begin{aligned}
\mathbb{P}(Y_1 \in E_1,\dots,Y_n \in E_n) &= \mathbb{P}(g_1(X_1) \in E_1,\dots,g_n(X_n) \in E_n) \\
&= \mathbb{P}(X_1 \in g_1^{-1}(E_1),\dots,X_n \in g_n^{-1}(E_n)) \\
&= \mathbb{P}(X_1 \in g_1^{-1}(E_1))\dots\mathbb{P}(X_n \in g_n^{-1}(E_n)) \\
&= \mathbb{P}(Y_1 \in E_1)\dots\mathbb{P}(Y_n \in E_n)
\end{aligned}$$

and the result follows. ∎

Definition 7.7. The joint characteristic function of n given random variables $X_1, \dots, X_n$ is the function $\varphi_{X_1,\dots,X_n} : \mathbb{R}^n \to \mathbb{C}$ given by

$$\varphi_{X_1,\dots,X_n}(t_1, \dots, t_n) = \mathbb{E}(e^{i(t_1 X_1 + \dots + t_n X_n)})$$

for $(t_1, \dots, t_n) \in \mathbb{R}^n$.

We state the following result, without proof.

Theorem 7.8. *The joint characteristic function uniquely determines the joint distribution function.*

Theorem 7.9. *Random variables $X_1, \dots, X_n$ are independent if and only if*

$$\mathbb{E}(g_1(X_1) \dots g_n(X_n)) = \mathbb{E}(g_1(X_1)) \dots \mathbb{E}(g_n(X_n)) \qquad (*)$$

for all Borel functions $g_1, \dots, g_n : \mathbb{R} \to \mathbb{C}$ such that each $g_1(X_1), \dots, g_n(X_n)$ is integrable.

Proof. Let $E_1, \dots, E_n$ be Borel sets in $\mathbb{R}$ and let $g_1 = \mathbb{1}_{E_1}, \dots, g_n = \mathbb{1}_{E_n}$. In this case, each $g_i : \mathbb{R} \to \mathbb{R}$ is a bounded Borel function and so $g_i(X_i)$ is integrable, $1 \le i \le n$.

Suppose that $(*)$ holds. Then

$$\mathbb{E}(\mathbb{1}_{E_1}(X_1) \dots \mathbb{1}_{E_n}(X_n)) = \mathbb{E}(\mathbb{1}_{E_1}(X_1)) \dots \mathbb{E}(\mathbb{1}_{E_n}(X_n)).$$

But $\mathbb{1}_{E_1}(X_1) \dots \mathbb{1}_{E_n}(X_n) : \Omega \to \mathbb{R}$ is the function which is equal to 1 if $X_i \in E_i$ for every $1 \le i \le n$ and is zero otherwise. Therefore its expectation is

$$\mathbb{E}(\mathbb{1}_{E_1}(X_1) \dots \mathbb{1}_{E_n}(X_n)) = 1.\mathbb{P}(X_1 \in E_1, \dots, X_n \in E_n).$$

Similarly, $\mathbb{E}(\mathbb{1}_{E_i}(X_i)) = 1.\mathbb{P}(X_i \in E_i)$ for $1 \le i \le n$. Therefore $(*)$ becomes

$$\mathbb{P}(X_1 \in E_1, \dots, X_n \in E_n) = \mathbb{P}(X_1 \in E_1) \dots \mathbb{P}(X_n \in E_n)$$

which means that $X_1, \dots, X_n$ are independent.

To prove the converse, suppose that $X_1, \dots, X_n$ are independent. Then, with the same notation as above, and working backwards, we find that

$$\begin{aligned}
\mathbb{E}(\mathbb{1}_{E_1}(X_1) \dots \mathbb{1}_{E_n}(X_n)) &= 1.\mathbb{P}(X_1 \in E_1, \dots, X_n \in E_n) \\
&= \mathbb{P}(X_1 \in E_1) \dots \mathbb{P}(X_n \in E_n), \quad \text{by independence,} \\
&= \mathbb{E}(\mathbb{1}_{E_1}(X_1)) \dots \mathbb{E}(\mathbb{1}_{E_n}(X_n)).
\end{aligned}$$

Thus $(*)$ holds when the g_i are indicator functions. By linearity, $(*)$ remains valid for linear combination of indicator functions and, in particular, if each g_i is a non-negative simple function on $\mathbb{R}$. The result now follows for non-negative g_i such that $g_i(X_i)$ is integrable by the Monotone Convergence Theorem. But then it holds for any g_i (such that $g_i(X_i)$ is integrable), again by linearity. ∎

Corollary 7.10. *The random variables $X_1, \dots, X_n$ are independent if and only if*

$$\varphi_{X_1,\dots,X_n}(t_1, \dots, t_n) = \varphi_{X_1}(t_1) \dots \varphi_{X_n}(t_n)$$

for all $(t_1, \dots, t_i) \in \mathbb{R}^n$.

Proof. If $X_1, \dots, X_n$ are independent, then the result is a direct consequence of Theorem 7.9.

Conversely, suppose that the joint characteristic function factorizes as above. Let $Y_1, \dots, Y_n$ be independent random variables such that $F_{X_i} = F_{Y_i}$ for all $1 \le i \le n$, i.e., Y_i has the same distribution as X_i. Then

$$\begin{aligned}
\varphi_{Y_1,\dots,Y_n}(t_1, \dots, t_n) &= \varphi_{Y_1}(t_1) \dots \varphi_{Y_n}(t_n), \quad \text{by independence,} \\
&= \varphi_{X_1}(t_1) \dots \varphi_{X_n}(t_n) \\
&= \varphi_{X_1,\dots,X_n}(t_1, \dots, t_n), \quad \text{by hypothesis.}
\end{aligned}$$

Since the joint distribution function is uniquely determined by the joint characteristic function, it follows that $F_{Y_1,\dots,Y_n} = F_{X_1,\dots,X_n}$. But $F_{Y_1,\dots,Y_n} = F_{Y_1} \dots F_{Y_n} = F_{X_1} \dots F_{X_n}$, by construction. Hence

$$F_{X_1,\dots,X_n} = F_{X_1} \dots F_{X_n}$$

which implies that $X_1, \dots, X_n$ are independent by Theorem 7.2. ∎

Chapter 8

Convergence of random variables.

We consider here some notions of convergence of random variables on a probability space $(\Omega, \mathcal{S}, \mathbb{P})$. We begin with a definition.

Definition 8.1. We say that random variables X and Y are equal almost surely (a. s.) if $\mathbb{P}(\{\, \omega \in \Omega : X(\omega) = Y(\omega)\,\}) = 1$.

Thus, the random variables X and Y are equal almost surely if they are equal with probability one.

It is convenient to observe at this point that the intersection of two events each with probability one also has probability one. Indeed, if A and B are events with $\mathbb{P}(A) = \mathbb{P}(B) = 1$, then $\mathbb{P}(A^c) = \mathbb{P}(B^c) = 0$ and so $0 \le \mathbb{P}((A \cap B)^c) = \mathbb{P}(A^c \cup B^c) \le \mathbb{P}(A^c) + \mathbb{P}(B^c) = 0$. Hence $\mathbb{P}(A \cap B) = 1$, as claimed.

Proposition 8.2. *Let $\sim$ be the relation defined by $X \sim Y$ if and only if $X = Y$ almost surely. Then $\sim$ is an equivalence relation on the collection of random variables on a probability space.*

Proof. It is clear that $X \sim X$ and also that $X \sim Y$ implies that $Y \sim X$.

Suppose that X, Y and V are random variables such that $X \sim Y$ and that $Y \sim V$. The proof is complete if we can show that $X \sim V$.

Let $A = \{\, \omega : X(\omega) = Y(\omega)\,\}$ and $B = \{\, \omega : Y(\omega) = V(\omega)\,\}$. By hypothesis, $\mathbb{P}(A) = \mathbb{P}(B) = 1$ and so, as noted above, $\mathbb{P}(A \cap B) = 1$. However, it is clear that $A \cap B \subseteq \{\, \omega : X(\omega) = V(\omega)\,\}$ and so we deduce that $\mathbb{P}(\{\, \omega : X(\omega) = V(\omega)\,\}) = 1$, i.e., $X = V$, a. s. ∎

Remark 8.3. In general, a property is said to hold almost surely (or with probability one) if the event that it fails has probability zero.

Proposition 8.4. *Suppose that $X \ge 0$ and $\mathbb{E}X = 0$. Then $X = 0$, a. s.*

Proof. For each $n \in \mathbb{N}$, let $A_n = \{\, \omega : X(\omega) > 1/n\,\}$. Then $A_n \in \mathcal{S}$, $A_n \subseteq A_{n+1}$ and

$$0 = \mathbb{E}X = \int_\Omega X \, d\mathbb{P}$$

$$\begin{aligned} &\geq \int_{A_n} X \, d\mathbb{P} \\ &\geq \tfrac{1}{n} \int_{A_n} d\mathbb{P} \\ &= \tfrac{1}{n} \, \mathbb{P}(A_n). \end{aligned}$$

Hence $\mathbb{P}(A_n) = 0$ for every $n \in \mathbb{N}$. But $\{\, \omega : X(\omega) > 0 \,\} = \bigcup_{n=1}^{\infty} A_n$ and so

$$\mathbb{P}(\{\, \omega : X(\omega) > 0 \,\}) = \lim_n \mathbb{P}(A_n) = 0$$

and so $\mathbb{P}(\{\, \omega : X(\omega) = 0 \,\}) = 1$. ∎

Corollary 8.5. *If* $\operatorname{var} X = 0$, *then* X *is almost surely constant.*

Proof. We have

$$0 = \operatorname{var} X = \mathbb{E}\, |X - \mathbb{E} X|^2 .$$

Since $|X - \mathbb{E} X|^2 \geq 0$, we conclude that $|X - \mathbb{E} X|^2 = 0$, a. s. and therefore $X = \mathbb{E} X$, a. s. ∎

Definition 8.6. The sequence (f_n) of random variables is said to converge almost surely to a random variable g if

$$\mathbb{P}(\{\, \omega : f_n(\omega) \to g(\omega), \text{ as } n \to \infty \,\}) = 1 \,,$$

i.e., $f_n(\omega) \to g(\omega)$, as $n \to \infty$, fails only for a collection of ωs forming a set of probability zero.

Remark 8.7. Put $h_n = |f_n - g|$. Then each h_n is a random variable. Set

$$A_m^k = \bigcap_{j > k} \{\, \omega : h_j(\omega) < 1/m \,\}.$$

Then $A_m^k \in \mathcal{S}$. Let $A_m = \bigcup_k A_m^k$. Then $A_m \in \mathcal{S}$ and

$$A_m = \{\, \omega : h_n(\omega) < 1/m, \;\text{ for all } n > \text{some } k \,\}.$$

Let $A = \bigcap_{m=1}^{\infty} A_m$. Then $A \in \mathcal{S}$ and

$$A = \{\, \omega : \text{for each } m \text{ there is } k \text{ such that } h_n(\omega) < 1/m, \text{ for all } n > k \,\}$$

(where k may depend on m). That is, $A = \{\, \omega : f_n(\omega) \to g(\omega), \text{ as } n \to \infty \,\}$. Since this set belongs to $\mathcal{S}$, it makes sense to ask whether $\mathbb{P}(A) = 1$ or not. In other words, the definition is meaningful.

Proposition 8.8. *If* $f_n \to f$, *a. s., as* $n \to \infty$ *and* $f_n \to g$, *a. s., as* $n \to \infty$, *then* $f = g$ *a. s.*

Proof. Let $A = \{\, \omega : f_n(\omega) \to f(\omega) \,\}$, $B = \{\, \omega : f_n(\omega) \to g(\omega) \,\}$ and $C = \{\, \omega : f(\omega) = g(\omega) \,\}$. By hypothesis, $\mathbb{P}(A) = 1$ and $\mathbb{P}(B) = 1$ and so $\mathbb{P}(A \cap B) = 1$. But $A \cap B \subseteq C$ which means that $\mathbb{P}(C) = 1$. ■

Proposition 8.9. *Suppose that $f_n \to f$, a.s. and $g_n \to g$, a.s. Then we have $f_n + g_n \to f + g$, a.s. and $f_n g_n \to fg$, a.s., as $n \to \infty$.*

Proof. Let $A = \{\, \omega : f_n(\omega) \to f(\omega) \,\}$ and $B = \{\, \omega : g_n(\omega) \to g(\omega) \,\}$. Then both $\mathbb{P}(A) = 1$ and $\mathbb{P}(B) = 1$. Hence $\mathbb{P}(A \cap B) = 1$. But each of the sets $\{\, \omega : f_n(\omega) + g_n(\omega) \to f(\omega) + g(\omega) \,\}$ and $\{\, \omega : f_n(\omega) g_n(\omega) \to f(\omega) g(\omega) \,\}$ contains $A \cap B$ and so each of these also has probability one. ■

Example 8.10. Let X have uniform distribution on $[0, 1]$. For each $n \in \mathbb{N}$, let $f_n = (\cos(2\pi X))^n$. Evidently, $f_n(\omega) \to 0$, as $n \to \infty$, *except* for those ω for which $X(\omega) = \frac{1}{2}k$, for some $k \in \mathbb{Z}$. But $\mathbb{P}(X = \frac{1}{2}k) = 0$ for every $k \in \mathbb{Z}$ and so $\mathbb{P}(X \neq 0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \dots) = 1$. We see that $f_n \to 0$, a.s., as $n \to \infty$.

Definition 8.11. We say that the sequence (f_n) of random variables converges in probability to the random variable f if for each $\varepsilon > 0$

$$\mathbb{P}(\, \{\, \omega : |f_n(\omega) - f(\omega)| \geq \varepsilon \,\} \,) \to 0$$

as $n \to \infty$.

Note that $f_n \to f$ in probability if and only if $(f_n - f) \to 0$ in probability.

Proposition 8.12. *If $f_n \to f$ and $g_n \to g$ in probability, then $f_n + g_n \to f + g$ in probability and $f_n g_n \to fg$ in probability.*

Proof. Let $\varepsilon > 0$ be given. First we note that

$$|f_n(\omega) + g_n(\omega) - f(\omega) - g(\omega)| \leq |f_n(\omega) - f(\omega)| + |g_n(\omega) - g(\omega)|$$

and so

$$\begin{aligned} \{\, \omega : |f_n(\omega) - f(\omega)| < \tfrac{1}{2}\varepsilon \,\} &\cap \{\, \omega : |g_n(\omega) - g(\omega)| < \tfrac{1}{2}\varepsilon \,\} \\ &\subseteq \{\, \omega : |f_n(\omega) + g_n(\omega) - f(\omega) - g(\omega)| < \varepsilon \,\}. \end{aligned}$$

Taking complements, we get

$$\begin{aligned} &\{\, \omega : |f_n(\omega) + g_n(\omega) - f(\omega) - g(\omega)| \geq \varepsilon \,\} \\ &\qquad \subseteq \{\, \omega : |f_n(\omega) - f(\omega)| \geq \tfrac{1}{2}\varepsilon \,\} \cup \{\, \omega : |g_n(\omega) - g(\omega)| \geq \tfrac{1}{2}\varepsilon \,\}. \end{aligned}$$

Hence

$$\begin{aligned} &\mathbb{P}(\{\, \omega : |f_n(\omega) + g_n(\omega) - f(\omega) - g(\omega)| \geq \varepsilon \,\}) \\ &\qquad \leq \mathbb{P}(\{\, \omega : |f_n(\omega) - f(\omega)| \geq \tfrac{1}{2}\varepsilon \,\}) + \mathbb{P}(\{\, \omega : |g_n(\omega) - g(\omega)| \geq \tfrac{1}{2}\varepsilon \,\}) \\ &\qquad \to 0\,, \end{aligned}$$

as $n \to \infty$.

For the last part, we first consider special cases. Suppose that $f_n \to 0$ in probability. We claim that $f_n g \to 0$ for any random variable g. (It is implicit that any random variable is finite-valued.) To see this, let $B_m = \{\, \omega : |g(\omega)| < m \,\}$, for each $m \in \mathbb{N}$. Then $B_m \subseteq B_{m+1}$ and $\bigcup_m B_m = \Omega$. Hence $\mathbb{P}(B_m) \uparrow 1$ and therefore $\mathbb{P}(\{\, \omega : |g(\omega)| \geq m \,\}) = \mathbb{P}(B_m^c) \to 0$, as $m \to \infty$. But for any given $\varepsilon > 0$, we have

$$\{\, \omega : |f_n(\omega) g(\omega)| \geq \varepsilon \,\} \subseteq \{\, \omega : |f_n(\omega)| \geq \varepsilon / m \,\} \cup \{\, \omega : |g(\omega)| \geq m \,\}$$

and so

$$\begin{aligned} &\mathbb{P}(\{\, \omega : |f_n(\omega) g(\omega)| \geq \varepsilon \,\}) \\ &\qquad \leq \mathbb{P}(\{\, \omega : |f_n(\omega)| \geq \varepsilon / m \,\}) + \mathbb{P}(\{\, \omega : |g(\omega)| \geq m \,\}) \qquad (*) \end{aligned}$$

for any $m \in \mathbb{N}$. For given $\delta > 0$, fix m so that $\mathbb{P}(\{\, \omega : |g(\omega)| \geq m \,\}) < \frac{1}{2}\delta$. By hypothesis, there is N such that $\mathbb{P}(\{\, \omega : |f_n(\omega)| \geq \varepsilon/m \,\}) < \frac{1}{2}\delta$ for all $n > N$ and so the right hand side of $(*)$ and therefore also the left hand side is $< \delta$ for all such n which means that $f_n g \to 0$ in probability, as claimed.

Next, we claim that if both $f_n \to 0$ and $g_n \to 0$ in probability, then $f_n g_n \to 0$ in probability. Indeed, for any $\varepsilon > 0$,

$$\{\, \omega : |f_n(\omega) g_n(\omega)| \geq \varepsilon \,\} \subseteq \{\, \omega : |f_n(\omega)| \geq \varepsilon \,\} \cup \{\, \omega : |g_n(\omega)| \geq 1 \,\}$$

and so

$$\begin{aligned} &\mathbb{P}(\{\, \omega : |f_n(\omega) g(\omega)| \geq \varepsilon \,\}) \\ &\qquad \leq \mathbb{P}(\{\, \omega : |f_n(\omega)| \geq \varepsilon \,\}) + \mathbb{P}(\{\, \omega : |g(\omega)| \geq 1 \,\}) \to 0\,, \end{aligned}$$

as $n \to \infty$.

We are now in a position to complete the proof. Suppose, then, that $f_n \to f$ and $g_n \to g$ in probability. We write

$$f_n g_n - f g = (f_n - f)(g_n - g) + (f_n - f) g + f (g_n - g)$$

and note that both $(f_n - f)$ and $(g_n - g)$ converge to 0 in probability, as $n \to \infty$. Using the results established above, we see that each of the three terms on the right hand side converges to 0 in probability and so, by the first part, does their sum. ∎

Proposition 8.13. *Suppose that $f_n \to f$ a. s. Then $f_n \to f$ in probability.*

Proof. Let $\varepsilon > 0$ and $\delta > 0$ be given. Set $g_n = |f_n - f|$ and let A be the set $A = \{\, \omega : g_n(\omega) \to 0, \ \text{ as } n \to \infty \,\}$. By hypothesis, $\mathbb{P}(A) = 1$.

Let us use the word 'eventually' to mean 'for sufficiently large n'. For $m \in \mathbb{N}$, let

$$A_m = \{\, \omega : g_n(\omega) < 1/m, \ \text{ eventually} \,\}.$$

Then $A_m = \bigcup_k A_m^k$ where

$$A_m^k = \{\, \omega : g_j(\omega) < 1/m, \text{ for all } j \geq k \,\} = \bigcap_{j>k} \{\, \omega : g_j(\omega) < 1/m \,\}.$$

Since $A \subseteq A_m$, it follows that $\mathbb{P}(A_m) = 1$. Furthermore, $A_m^k \subseteq A_m^{k+1}$ and so (since $A_m = \bigcup_k A_m^k$) we have $\mathbb{P}(A_m^k) \uparrow 1$ as $k \to \infty$.

Fix $m \in \mathbb{N}$ such that $1/m < \varepsilon$. Then $\mathbb{P}(A_m^k) > 1 - \delta$ for large k, that is, $\mathbb{P}(\{\, \omega : g_j(\omega) < 1/m, \text{ for all } j > k \,\}) > 1 - \delta$, for all sufficiently large k.

Hence, for any $j > k$, the fact that

$$\{\, \omega : g_j(\omega) < 1/m \,\} \supseteq \{\, \omega : g_j(\omega) < 1/m, \text{ for all } j \geq k \,\}$$

implies that

$$\mathbb{P}(\{\, \omega : g_j(\omega) < 1/m \,\}) \geq \mathbb{P}(\{\, \omega : g_j(\omega) < 1/m, \text{ for all } j \geq k \,\}) > 1 - \delta$$

for sufficiently large k. Finally, we have

$$\mathbb{P}(\{\, \omega : g_j(\omega) \geq \varepsilon \,\}) \leq \mathbb{P}(\{\, \omega : g_j(\omega) \geq 1/m \,\}) < \delta$$

for suitably large k and all $j > k$. ■

The converse to this is false as the next example shows.

Example 8.14. Let X be a random variable with uniform distribution on the interval $[0,1]$. For each $m \in \mathbb{N}$, construct the m subintervals $J_1^m = [0, 1/m]$, $J_2^m = [1/m, 2/m]$, ..., $J_m^m = [(m-1)/m, 1/m]$ of $[0,1]$ and let (I_n) be the sequence $I_1 = J_1^1$, $I_2 = J_1^2$, $I_3 = J_2^2$, $I_4 = J_1^3$, $I_5 = J_2^3 \ldots$. The point is that as n increases, the intervals I_n become narrower and narrower but they nevertheless continue to step across the whole interval $[0,1]$. In particular, every point $a \in [0,1]$ belongs to infinitely-many I_ns.

Let $f_n = \mathbb{1}_{I_n}(X)$. We claim that $f_n \to 0$ in probability. To see this, let $\varepsilon > 0$ and $\delta > 0$ be given. Now, by definition, $f_n(\omega)$ is equal to either 0 or 1 and so $\mathbb{P}(\{\, \omega : |f_n(\omega)| \geq \varepsilon \,\}) = 0$ if $\varepsilon > 1$. On the other hand, for any $\varepsilon \leq 1$,

$$\begin{aligned}
\mathbb{P}(\{\, \omega : |f_n(\omega)| \geq \varepsilon \,\}) &= \mathbb{P}(\{\, \omega : |f_n(\omega)| = 1 \,\}) \\
&= \mathbb{P}(\{\, \omega : X(\omega) \in I_n \,\}) \\
&= \text{length}(I_n) \\
&< \delta
\end{aligned}$$

for all sufficiently large n. Hence $f_n \to 0$ in probability.

However, let $A = \{\, \omega : X(\omega) \in [0,1] \,\}$. Then $\mathbb{P}(A) = 1$ since X has uniform distribution on $[0,1]$. For any $\omega \in A$, $X(\omega) \in [0,1]$ and so $X(\omega) \in I_k$ for infinitely-many ks. For such k, $f_k(\omega) = \mathbb{1}_{I_k}(X(\omega)) = 1$ and so the sequence $(f_n(\omega))$ does not converge to 0. Therefore $\{\, \omega : f_n(\omega) \to 0 \,\} \subseteq A^c$ and so we see that $\mathbb{P}(\{\, \omega : f_n(\omega) \to 0 \,\}) = 0 \neq 1$ which means that it is *false* that $f_n \to 0$ a.s.

Chapter 9

The strong law of large numbers.

First, we recall the weak law of large numbers.

Proposition 9.1 (Chebyshev's inequality). *For any non-negative random variable Y and any $b > 0$, we have*

$$\mathbb{P}(Y \geq b) \leq \frac{\mathbb{E}Y}{b}\,.$$

Proof. Let $A = \{\,\omega : Y(\omega < b\,\}$ and let $B = \{\,\omega : Y(\omega) \geq b\,\}$. Evidently, $A \cup B = \Omega$ and $A \cup B = \varnothing$. Hence

$$\begin{aligned}
\mathbb{E}(Y) = \mathbb{E}(\mathbb{1}_A Y + \mathbb{1}_B Y) &= \int_\Omega \mathbb{1}_A Y \, d\mathbb{P} + \int_\Omega \mathbb{1}_B Y \, d\mathbb{P} \\
&\geq \int_\Omega \mathbb{1}_B Y \, d\mathbb{P} = \mathbb{E}(\mathbb{1}_B Y), \quad \text{since } Y \geq 0, \\
&\geq b \int_\Omega \mathbb{1}_B \, d\mathbb{P} = \mathbb{E}(\mathbb{1}_B b), \quad \text{since } Y \geq b \text{ on } B, \\
&= b\,\mathbb{P}(B)\,,
\end{aligned}$$

as required. ■

Corollary 9.2. *Let X be any random variable. Then, for any $c \in \mathbb{R}$, $b > 0$ and $m > 0$, we have*

$$\mathbb{P}(|X - c| \geq b) \leq \frac{\mathbb{E}(|X - c|^m)}{b^m}\,.$$

In particular, if $\mathbb{E}X = \mu$ and $\operatorname{var} X = \sigma^2$, then

$$\mathbb{P}(|X - \mu| \geq b) \leq \frac{\sigma^2}{b^2}\,.$$

Proof. Set $Y = |X - c|^m$. Then $Y \geq 0$ and so

$$\mathbb{P}(|X - c| \geq b) = \mathbb{P}(|X - c|^m \geq b^m) \leq \frac{\mathbb{E}(|X - c|^m)}{b^m}$$

by Chebyshev's inequality, Proposition 9.1. Putting $c = \mu$ and $m = 2$ gives

$$\mathbb{P}(|X - \mu| \geq b) \leq \frac{\operatorname{var} X}{b^2} = \frac{\sigma^2}{b^2},$$

as claimed. ∎

Theorem 9.3 (Weak law of large numbers). *Let $X_1, X_2, \dots$ be a sequence of independent random variables and suppose that there is some constant $M > 0$ such that $E(X_i^2) \leq M$ for all $i = 1, 2, \dots$. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\Big(\Big| \frac{X_1 + X_2 + \cdots + X_n - (\mathbb{E} X_1 + \cdots + \mathbb{E} X_n)}{n} \Big| > \varepsilon \Big) \to 0$$

as $n \to \infty$.

Proof. First note that by Schwarz' inequality,

$$\mathbb{E}(|X_i|)^2 = \mathbb{E}(|X_i|\,1)^2 \leq \mathbb{E}(|X_i|^2) \leq M^2$$

and therefore $\operatorname{var} X_i = \mathbb{E}(X_i^2) - (\mathbb{E} X_i)^2 \leq 2M^2$ for all $i \in \mathbb{N}$.

Set $S_n = X_1 + \cdots + X_n$. Then, for any given $\varepsilon > 0$, Chebyshev's inequality implies that

$$\begin{aligned}
\mathbb{P}\Big(\Big| \frac{S_n - \mathbb{E} S_n}{n} \Big| \geq \varepsilon \Big) &\leq \frac{\mathbb{E}\big(\big(\frac{S_n - \mathbb{E} S_n}{n} \big)^2 \big)}{\varepsilon^2} \\
&= \frac{\operatorname{var}(S_n/n)}{\varepsilon^2} \\
&= \frac{\frac{1}{n^2} \operatorname{var}(X_1 + \cdots + X_n)}{\varepsilon^2} \\
&= \frac{\operatorname{var} X_1 + \cdots + \operatorname{var} X_n}{n^2 \varepsilon^2}, \quad \text{by independence,} \\
&\leq \frac{n 2M^2}{n^2 \varepsilon^2} = \frac{2M^2}{n \varepsilon^2} \\
&\to 0,
\end{aligned}$$

as $n \to \infty$. ∎

Remark 9.4. The weak law of large numbers says that the random variables $\frac{1}{n}(S_n - \mathbb{E} S_n)$ converge to 0 in probability. If $\mathbb{E} X_i = \mu$, for all i, then this becomes the statement that the $\frac{1}{n} S_n \to \mu$ in probability.

Theorem 9.5 (Borel-Cantelli lemma). *Suppose that $A_1, A_2, \dots$ is a sequence of events such that the series $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converges. Then*

$$\mathbb{P}(\{ \omega : \omega \in A_n \text{ for infinitely-many } n \}) = 0.$$

Proof. Let $G = \{\, \omega : \omega \in A_n$ for infinitely-many $n \,\}$. Then we see that

$$G = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \,.$$

Hence

$$\begin{aligned}
\mathbb{P}(G) &= \mathbb{P}\Big(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \Big) \\
&\le \mathbb{P}\Big(\bigcup_{k=n}^{\infty} A_k \Big), \quad \text{for any } n, \\
&\le \sum_{k=n}^{\infty} \mathbb{P}(A_k) \\
&\to 0
\end{aligned}$$

as $n \to \infty$, since the series $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converges. It follows that the left hand side is zero. ∎

We shall now prove (a version of) the strong law of large numbers.

Theorem 9.6 (Strong law of large numbers). *Let $X_1, X_2, \dots$ be a sequence of independent random variables and suppose that there is some constant $M > 0$ such that $\mathbb{E}((X_i - \mathbb{E}X_i)^4) < M$ for all $i \in \mathbb{N}$. Let $S_n = X_1 + \dots + X_n$. Then*

$$\frac{S_n - \mathbb{E}(S_n)}{n} \to 0 \quad \text{a. s.}$$

as $n \to \infty$. In particular, if $\mathbb{E}X_i = \mu$ for all i, then $\mathbb{E}S_n = n\mu$ and so it follows that $\frac{1}{n} S_n \to \mu$ a. s.

Proof. Firstly, setting $Y_i = X_i - \mathbb{E}X_i$, we see that $\mathbb{E}Y_i = 0$, $\mathbb{E}(Y^4) < M$ and $S_n - \mathbb{E}(S_n) = Y_1 + \dots + Y_n$. So by considering Y_i instead of X_i, we may assume that $\mathbb{E}X_i = 0$.

The idea of the proof is to use the Borel-Cantelli Lemma and Chebyshev's inequality. To this end, we shall show that $\sum_{n=1}^{\infty} \mathbb{E}((\frac{1}{n} S_n)^4)$ converges.

We have

$$\begin{aligned}
S_n^4 &= (X_1 + \dots + X_n)^4 \\
&= \sum_{i=1}^{n} X_i^4 + \sum_{j<k} \binom{4}{2} X_i^2 X_j^2 + \sum_{i,k,\ell \text{ distinct}} 12\, X_i^2 X_k X_\ell \\
&\qquad + \sum_{j,k,\ell,m \text{ distinct}} X_j X_k X_\ell X_m + \sum_{j \ne k} 12\, X_j^3 X_k .
\end{aligned}$$

But, by independence,

$$\mathbb{E}(X_i^2 X_k X_\ell) = \mathbb{E}(X_i^2) \mathbb{E}(X_k) \mathbb{E}(X_\ell) = 0.$$

Similarly, both $\mathbb{E}(X_j X_k X_\ell X_m) = 0$ and $\mathbb{E}(X_j^3 X_k) = 0$. Hence

$$\mathbb{E}(S_n^4) = \sum_{i=1}^{n} \mathbb{E}(X_i^4) + \sum_{j<k} 6\, \mathbb{E}(X_i^2)\, \mathbb{E}(X_j^2).$$

By Schwarz' inequality,

$$\mathbb{E}(X_j^2) = \mathbb{E}(X_j^2\, 1) \le \mathbb{E}(X_j^4)^{1/2}\, \mathbb{E}(1^2)^{1/2} \le M^{1/2}.$$

This gives

$$\mathbb{E}(S_n^4) \le nM + 6 \tbinom{n}{2} M = (3n^2 - 2n)M < 3n^2 M$$

and so

$$\sum_{n=1}^{\infty} \mathbb{E}((\tfrac{1}{n} S_n)^4) < \sum_{n=1}^{\infty} \frac{3n^2 M}{n^4} = \sum_{n=1}^{\infty} \frac{3M}{n^2}$$

which converges.

Now, for any $m \in \mathbb{N}$, Chebyshev's inequality gives

$$\mathbb{P}(\big|\tfrac{1}{n} S_n\big| \ge \tfrac{1}{m}) \le \mathbb{E}((\tfrac{1}{n} S_n)^4)/(\tfrac{1}{m})^4$$

and so it follows from the discussion above that $\sum_{n=1}^{\infty} \mathbb{P}(|\frac{1}{n} S_n| \ge \frac{1}{m})$ is convergent. By the Borel-Cantelli Lemma, it follows that $\mathbb{P}(G_m) = 0$, where G_m is the event

$$G_m = \{\, \omega : |\tfrac{1}{n} S_n| \ge \tfrac{1}{m} \text{ for infinitely-many } n \,\}$$

and therefore $\mathbb{P}\big(\bigcup_{m=1}^{\infty} G_m\big) = 0$. But to say that ω is such that $\frac{1}{n} S_n(\omega) \not\to 0$ is to say that there is some $m_0 > 0$ such that $|\frac{1}{n} S_n(\omega)| \ge 1/m_0$ for infinitely-many n. In other words,

$$\{\, \omega : \tfrac{1}{n} S_n(\omega) \not\to 0 \,\} = \bigcup_{m=1}^{\infty} G_m$$

and we conclude that $\mathbb{P}(\{\, \omega : \frac{1}{n} S_n(\omega) \not\to 0 \,\}) = 0$ which means that $\frac{1}{n} S_n \to 0$ a. s. and the proof is complete. ∎

Remark 9.7. The condition on the expectations of the fourth powers can be relaxed — but then the proof is harder.

Chapter 10

Stochastic processes.

A stochastic process is simply a "labelled" family of random variables on a probability space. The label is usually motivated by physical applications is considered to represent "time".

Examples 10.1.

1. For $n = 1, 2, 3, \dots$, X_n could represent the number of customers entering a shop on day n.

2. X_n, $n \in \mathbb{N}$, could represent the bacteria population in a petri dish after n minutes.

3. X_n, $n \in \mathbb{N}$, is the number of heads after n tosses of a coin.

4. X_t, $t \geq 0$, could represent the number of particles emitted by a radioactive source after time t.

5. For each $t \geq 0$, X_t could be a random variable with a Poisson distribution with mean μt. (This is called a Poisson process and is a good model for the number of particles emitted in radioactive decay.)

6. X_t, $t \geq 0$, could represent the position of a pollen particle at time t when subjected to random collisions with air molecules. This "erratic" motion was observed by the botanist Brown and later discussed by Einstein. The mathematics behind this so-called "Brownian motion" was developed by Norbert Wiener (so it is also referred to as a Wiener process).

Let $\{ X_\alpha : \alpha \in \Lambda \}$ be a stochastic process indexed by $\alpha \in \Lambda$. As noted above, Λ is just some given set of labels, but in practice one usually takes $\Lambda = \mathbb{N}$ or $\Lambda = [0, \infty)$ (or $[0, T]$) or something of this kind.

For $\alpha_1, \dots, \alpha_n \in \Lambda$,

$$F_{\alpha_1,\dots,\alpha_n}(x_1, \dots, x_n) \equiv \mathbb{P}(X_{\alpha_1} \leq x_1, \dots, X_{\alpha_n} \leq x_n)$$

is the joint distribution function of $X_{\alpha_1}, \dots, X_{\alpha_n}$. These are increasing, right-continuous functions of $x_1, \dots, x_n$ for any choice of n and choice of labels $\alpha_1, \dots, \alpha_n \in \Lambda$. Furthermore,

$$F_{\alpha_1,\dots,\alpha_n}(x_1, \dots, x_n) \to F_{\alpha_1,\dots,\widehat{\alpha_j},\dots,\alpha_n}(x_1, \dots, \widehat{x_j}, \dots, x_n)$$

as $x_j \to \infty$, $1 \le j \le n$. That is, by letting $x_j \to \infty$, the distribution function $F_{\alpha_1,\dots,\alpha_n}(x_1, \dots, x_n)$ "reduces" to $F_{\alpha_1,\dots,\widehat{\alpha_j},\dots,\alpha_n}(x_1, \dots, \widehat{x_j}, \dots, x_n)$. This is called a *consistency condition*.

One of the fundamental results in the theory of stochastic processes is the theorem of Daniel-Kolmogorov which says that given *any* collection of such functions $F_{\alpha_1,\dots,\alpha_n}(x_1, \dots, x_n)$ indexed by finite subsets of Λ and satisfying the consistency conditions, there exists a probability space $(\Omega, \mathcal{S}, \mathbb{P})$ and a stochastic process $\{ X_\alpha : \alpha \in \Lambda \}$, labelled by Λ, on $(\Omega, \mathcal{S}, \mathbb{P})$ such that the given $F_{\alpha_1,\dots,\alpha_n}(x_1, \dots, x_n)$s are the joint distribution functions for the process. (This is a generalization of the theorem that any increasing, right-continuous . . . etc function actually *is* the distribution function of some random variable.)

The idea of the proof is to explicitly construct Ω and $\mathbb{P}$. One takes Ω to be the collection of all functions from Λ into $\mathbb{R}$ and then, for $\alpha \in \Lambda$, defines X_α to be "evaluation at α":

$$X_\alpha(\varphi) = \varphi(\alpha), \quad \varphi \in \Omega.$$

The given $F_{\alpha_1,\dots}$s are then used to construct a probability measure on a suitable σ-algebra of subsets of Ω. The details are rather technical, as one might guess.

Suppose that $\{ X_n : n = 0, 1, 2, \dots \}$ is a stochastic process (indexed by $\{ 0 \} \cup \mathbb{N}$). Then it is called a Markov process if, for any n, the probabilities of events concerning the X_n for $n > m$ given information about the X_n for $n \le m$ is the same as given only the information about X_m. For example, if $\{ X_n \}$ is a Markov process, then, for events A_1, A_2, A_3,

$$\mathbb{P}(\, X_{10} \in A_1 \,|\, X_8 \in A_2,\, X_9 \in A_3 \,) = \mathbb{P}(\, X_{10} \in A_1 \,|\, X_9 \in A_3 \,).$$

A Markov process takes account of the present, but forgets the past. It has no memory. There is a similar definition for Markov processes indexed by $[0, \infty)$.

Simple one-dimensional random walk.

We wish to discuss Brownian motion — a dust particle bombarded by air molecules. To simplify matters, we consider motion in just one dimension.

Suppose, then, that we have a random walk — after each time interval Δt, a particle moves one step to the right or one step to the left with equal probability . Let Δd denote the size of the step. We shall consider both Δt and Δd to be very small — so that there are very many but very small jumps as time goes by.

Let $Y_1, Y_2, \dots$ be a sequence of independent random variables such that $\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = \frac{1}{2}$. Thus, $\mathbb{E}\, Y_i = 0$ and $\operatorname{var} Y_i = \frac{1}{2} + \frac{1}{2} = 1$. After n jumps, the position of the random walk, *given that it started at* 0, is just

$$X^{(n)} = \Delta d\, Y_1 + \Delta d\, Y_2 + \cdots + \Delta d\, Y_n \,.$$

Given n, suppose that $\Delta t = 1/2^n$. Let $t = k/2^n$ and let $X_t^{(n)}$ be the position of the particle at time t, i.e., after k jumps:

$$X_t^{(n)} = \Delta d\, Y_1 + \Delta d\, Y_2 + \cdots + \Delta d\, Y_k \,.$$

Consider t to be fixed but with n increasing, which means more and more jumps, $k = 2^n t$. So

$$X_t^{(n)} = \Delta d\, Y_1 + \Delta d\, Y_2 + \cdots + \Delta d\, Y_{2^n t} \,.$$

Now, $\operatorname{var}(\Delta d\, Y_i) = (\Delta d)^2$ and $\mathbb{E}(\Delta d\, Y_i) = 0$ and so, by the Central Limit Theorem,

$$\mathbb{P}\Big(\frac{X_t^{(n)}}{\sqrt{2^n t}\, \Delta d} \le a \Big) = \mathbb{P}\Big(\frac{\Delta d\, Y_1 + \Delta d\, Y_2 + \cdots + \Delta d\, Y_{2^n t}}{\sqrt{2^n t}\, \Delta d} \le a \Big) \to \Phi(a)$$

as $n \to \infty$, i.e., $\frac{X_t^{(n)}}{\sqrt{2^n t}\, \Delta d}$ is approximately a standard normal random variable.

But suppose we also let $\Delta d \to 0$, as $n \to \infty$, in such a way that $\Delta d/\sqrt{\Delta t} = 1$, i.e., $\Delta d/\sqrt{1/2^n} = 1$, i.e., $\Delta d \sqrt{2^n} = 1$. Then it follows that $\frac{X_t^{(n)}}{\sqrt{2^n t}\, \Delta d} = \frac{X_t}{\sqrt{t}}$ is approximately a standard normal random variable, i.e., in the limit $n \to \infty$, X_t, the position of the particle after time t, has a *normal distribution with mean* 0 *and variance* t.

Consider now, two times, $s = m/2^n$ and $t = k/2^n$, with $t > s$ (i.e., $k > m$). Then

$$\begin{aligned} X_t^{(n)} - X_s^{(n)} &= \Delta d\, Y_1 + \cdots + \Delta d\, Y_1 - \Delta d\, Y_1 - \cdots - \Delta d\, Y_m \\ &= \Delta d\, Y_{m+1} + \cdots + \Delta d\, Y_k. \end{aligned}$$

As before, we find that

$$\frac{X_t^{(n)} - X_s^{(n)}}{\sqrt{(k-m)}\, \Delta d} = \frac{X_t^{(n)} - X_s^{(n)}}{\sqrt{(t-s) 2^n}\, \Delta d} \to N(0,1)$$

as $n \to \infty$, i.e., $\frac{X_t^{(n)} - X_s^{(n)}}{\sqrt{(t-s)}}$ is approximately standard normal and so $X_t - X_s$ is approximately $N(0, t-s)$ (normal, mean 0, variance $t - s$).

Note also that $X_t^{(n)} - X_s^{(n)}$ depends only on Y_i for $i > m$ and $X_s^{(n)}$ depends on Y_j for $i \le m$, so these are independent.

For given intervals (α, β) and (a, b), we have

$$\begin{aligned}
&\mathrm{Prob}(X_t^{(n)} \in (\alpha,\beta), X_s^{(n)} \in (a,b)) \\
&\quad = \sum_{x' \in (a,b)} \mathrm{Prob}(X_t^{(n)} \in (\alpha,\beta) | X_s^{(n)} = x')\,\mathrm{Prob}(X_s^{(n)} = x') \\
&\quad = \sum_{x' \in (a,b)} \mathrm{Prob}(X_{k-m \text{ steps}}^{(n)} \in (\alpha,\beta) - x')\,\mathrm{Prob}(X_s^{(n)} = x') \\
&\quad = \sum_{a < x' < b} \mathrm{Prob}(X_{t-s}^{(n)} + x' \in (\alpha,\beta))\,\mathrm{Prob}(X_s^{(n)} = x').
\end{aligned}$$

For large n, $X_{t-s}^{(n)} + x'$ is approximately $N(x', (t-s))$ and so

$$\mathrm{Prob}(X_{t-s}^{(n)} + x' \in (\alpha,\beta)) \simeq \int_\alpha^\beta \frac{e^{-\frac{1}{2}(x-x')^2/(t-s)}}{\sqrt{2\pi(t-s)}}\,dx.$$

Furthermore,

$$\mathrm{Prob}(X_s^{(n)} = x') = \mathrm{Prob}(x' \le X_s^{(n)} < x' + \Delta d) \simeq \varphi_{X_s}(x')\,\Delta d$$

where $\varphi_{X_s}(x') = e^{-\frac{1}{2}x'^2/s}/\sqrt{2\pi s}$ is the $N(0,s)$ density. The sum is a Riemann-sum approximation to the Riemann integral, so letting $n \to \infty$, we obtain the formula

$$\begin{aligned}
&\mathrm{Prob}(X_t \in (\alpha,\beta), X_s \in (a,b)) \\
&\qquad\qquad = \int_a^b dx' \int_\alpha^\beta dx\, \frac{e^{-(x-x')^2/2(t-s)}}{\sqrt{2\pi(t-s)}}\, \frac{e^{-x'^2/2s}}{\sqrt{2\pi s}}\,.
\end{aligned}$$

Using this method, we can find the joint distribution functions $F_{t_1,\dots,t_n}$ for the process $\{\, X_t : t \ge 0 \,\}$ — this is one-dimensional Brownian motion.

There is a connection with the heat equation, as follows. Set

$$p(x;t) = \tfrac{1}{\sqrt{2\pi t}}\,\exp\bigl(-\tfrac{x^2}{2t}\bigr).$$

Then one sees that $2\frac{\partial p}{\partial t} = \frac{\partial^2 p}{\partial x^2}$, i.e., $p(x;t)$ satisfies the one-dimensional heat equation. (Note that the change of variable $t = 2\tau$ gives $\frac{\partial p}{\partial \tau} = \frac{\partial^2 p}{\partial x^2}$.)

For given $f : \mathbb{R} \to \mathbb{R}$, define

$$\varphi(x,t) = \int_{-\infty}^{\infty} f(y)\,p(x-y;t)\,dy.$$

Then (assuming f is sufficiently well-behaved),

$$2\,\frac{\partial\varphi}{\partial t} = \frac{\partial^2\varphi}{\partial^2 x}\,,$$

i.e., φ satisfies the heat equation. Furthermore, as $t \downarrow 0$, $\varphi(x:t) \to f(x)$. To see this, we write

$$\begin{aligned}\varphi(x;t) &= \int_{-\infty}^{\infty} f(y)\, p(x-y:t)\, dy \\ &= \int_{-\infty}^{\infty} f(x-z)\, p(z:t)\, dz\,, \quad \text{setting } z = x-y, \\ &= f(x) + \int_{-\infty}^{\infty} \big(f(x-z) - f(x)\big)\, p(z:t)\, dz\,,\end{aligned}$$

since $\int_{-\infty}^{\infty} p(z:t)\, dz = 1$. Now,

$$\begin{aligned}\int_{-\infty}^{\infty} \big(f(x-z) - f(x)\big)\, p(z:t)\, dz &= \int_{-\infty}^{\infty} \big(f(x-z) - f(x)\big)\, \frac{e^{-z^2/2t}}{\sqrt{2\pi t}}\, dz \\ &= \int_{-\infty}^{\infty} \big(f(x-w\sqrt{t}) - f(x)\big)\, \frac{e^{-w^2/2}}{\sqrt{2\pi}}\, dw\,,\end{aligned}$$

using the change of variable $w = z/\sqrt{t}$. The integrand $\to 0$ as $t \downarrow 0$ which establishes our claim. (It is enough for f to be continuous and bounded.)

We have established that

$$\varphi(x,t) = \int_{-\infty}^{\infty} f(y)\, p(x-y;t)\, dy$$

satisfies the heat equation and the initial condition $\lim_{t\downarrow 0} \varphi(x;t) = f(x)$.

Now let $X_t^{(x)}$ be Brownian motion starting from x. Then $X_t^{(x)} = X_t + x$, where X_t is Brownian motion starting from 0. (To be precise, we should really say that $X_t^{(x)}$ and $X_t + x$ have the same distribution, rather than that they are equal.) $X_t^{(x)}$ has a $N(x,t)$ distribution and so

$$\mathbb{E}(f(X_t^{(x)})) = \int_{-\infty}^{\infty} f(y)\, \frac{e^{-(y-x)^2/2t}}{\sqrt{2\pi t}}\, dy = \varphi(x;t)$$

as above. Hence $\mathbb{E}(f(X_t^{(x)}))$ satisfies the heat equation (also called the diffusion equation) with initial condition $\mathbb{E}(f(X_t^{(x)})) \to f(x)$ as $t \downarrow 0$.

This connection between Brownian motion and the heat equation has proved very fruitful.