

1 Measures

Let E be any set. A collection \mathcal{E} of subsets of E is called a σ -algebra if the following holds:

1. $\emptyset \in \mathcal{E}$.
2. If $A \in \mathcal{E}$, then $A^c = E \setminus A \in \mathcal{E}$.
3. If $(A_n : n \in \mathbb{N})$, $A_n \in \mathcal{E}$, then $\bigcup_n A_n \in \mathcal{E}$.

Examples.

- $\mathcal{E} = \{\emptyset, E\}$
- $\mathcal{E} = \mathcal{P}(E)$, the set of all subsets of E .

Note that $\bigcap_n A_n = (\bigcup_n A_n^c)^c$, any σ -algebra \mathcal{E} is also closed under countable intersection of its elements. Also $B \setminus A = B \cap A^c \in \mathcal{E}$ whenever $A, B \in \mathcal{E}$.

Any set E with a choice of σ -algebra \mathcal{E} is called a *measurable space*, and the elements of \mathcal{E} are called *measurable sets*.

A *measure* μ is a set-function $\mu : \mathcal{E} \rightarrow [0, \infty]$ such that $\mu(\emptyset) = 0$, and for any $(A_n : n \in \mathbb{N})$, $A_n \in \mathcal{E}$ pairwise disjoint ($A_n \cap A_m = \emptyset$ for all $n \neq m$) then

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n) \quad (\text{countable additivity of } \mu)$$

If \mathcal{E} is countable, then for any $A \in \mathcal{P}(E)$ and a measure μ

$$\mu(A) = \mu\left(\bigcup_{x \in A} \{x\}\right) = \sum_{x \in A} \mu(\{x\})$$

hence there is a one-to-one correspondence between measures and mass functions on E .

For any collection \mathcal{A} of subsets of E , we define the σ -algebra $\sigma(\mathcal{A})$ generated by \mathcal{A} as

$$\sigma(\mathcal{A}) = \{A \subseteq E : A \in \mathcal{E} \text{ } \forall \sigma\text{-algebras } \mathcal{E} \supseteq \mathcal{A}\}$$

which equals (Example sheet)

$$\sigma(\mathcal{A}) = \bigcap_{\sigma\text{-algebras } \mathcal{E} \supseteq \mathcal{A}} \mathcal{E}$$

To construct good ‘generators’ we define

1. \mathcal{A} is called a *ring over E* if $\emptyset \in \mathcal{A}$; if $A, B \in \mathcal{A}$ then $B \setminus A \in \mathcal{A}$ and $A \cup B \in \mathcal{A}$.

2. \mathcal{A} is called an *algebra over E* if $\emptyset \in \mathcal{A}$; if $A, B \in \mathcal{A}$ then $A^c \in \mathcal{A}$, $A \cup B \in \mathcal{A}$.

Notice that in a ring $A \Delta B = (B \setminus A) \cup (A \setminus B) \in \mathcal{A}$ and $A \cap B = (A \cup B) \setminus (A \Delta B) \in \mathcal{A}$. Also, $B \setminus A = B \cap A^c = (B^c \cup A)^c \in \mathcal{A}$, so an algebra is a ring.

Fact: If $\bigcup_n A_n$, $A_n \in \mathcal{E}$, \mathcal{E} some σ -algebra (or a ring if the union is finite) - then we can find $B_n \in \mathcal{E}$ disjoint such that $\bigcup_n A_n = \bigcup_n B_n$. Indeed, define $\tilde{A}_n = \bigcup_{j \leq n} A_j$, and set $B_n = \tilde{A}_n \setminus \tilde{A}_{n-1}$, then the fact follows. [“disjointification of countable unions”]

Definition. A *set function* on any collection \mathcal{A} of subsets of E (where $\emptyset \in \mathcal{A}$) is a map $\mu : \mathcal{A} \rightarrow [0, \infty]$ such that $\mu(\emptyset) = 0$. We say μ is

1. *increasing* if $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$; $A, B \in \mathcal{A}$
2. *additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \in \mathcal{A}$; $A \cup B \in \mathcal{A}$ and $A \cap B = \emptyset$.
3. *countably additive* if $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ for any $(A_n : n \in \mathbb{N})$ where $A_n \in \mathcal{A}$ disjoint and $\bigcup_n A_n \in \mathcal{A}$.
4. *countably sub-additive* if $\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$ for all $(A_n : n \in \mathbb{N})$ such that $\bigcup_n A_n \in \mathcal{A}$

Remark: one can show that a measure μ on a σ -algebra satisfies 1-4 above.

Theorem (Caratheodory). *Let μ be a countably additive set function on a ring \mathcal{A} of subsets of E . Then there exists a measure μ^* on $\sigma(\mathcal{A})$ such that $\mu^*|_{\mathcal{A}} = \mu$.*

Proof. For $B \subseteq E$ define the *outer measure* μ^* as

$$\mu^*(B) = \inf \left\{ \sum_{n \in \mathbb{N}} \mu(A_n) : A_n \in \mathcal{A}, B \subseteq \bigcup_n A_n \right\}$$

and set $\mu^*(B) = \infty$ if the set within the infimum is empty.

Define

$$\mathcal{M} = \{A \subseteq E : \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c) \forall B \subseteq E\}$$

the “ μ^* -measurable” sets.

Step 1: μ^* is countably sub-additive on $\mathcal{P}(E)$. For any $B \subseteq E$ and $B_n \subseteq E$ such that $B \subseteq \bigcup_n B_n$ we have

$$\mu^*(B) \leq \sum_n \mu^*(B_n) \quad (\dagger)$$

WLOG we assume $\mu^*(B_n) < \infty$ for all n so for all $\varepsilon > 0$, there exists A_{nm} such that $B_n \subseteq \bigcup_m A_{nm}$ and

$$\mu^*(B_n) + \frac{\varepsilon}{2^n} \geq \sum_m \mu(A_{nm})$$

Now since μ^* and since $B \subseteq \bigcup_n B_n \subseteq \bigcup_{n,m} A_{nm}$, hence

$$\mu^*(B) \leq \mu^*\left(\bigcup_{n,m} A_{nm}\right) \leq \sum_{n,m} \mu(A_{nm}) \leq \sum_n \mu^*(B_n) + \underbrace{\sum_n \frac{\varepsilon}{2^n}}_{=\varepsilon}$$

so (\dagger) follows since ε was arbitrary.

Step 2: μ^* extends μ . Let $A \in \mathcal{A}$. Clearly $A = A \cup \emptyset \cup \dots \cup \emptyset$, so by definition of μ^* , $\mu^*(A) \leq \mu(A) + 0 + \dots + 0$. So we need to prove $\mu(A) \leq \mu^*(A)$. Again, assume $\mu^*(A) < \infty$ WLOG, and let $A_n \in \mathcal{A}$ be such that $A \subseteq \bigcup_n A_n$. Then $A = \bigcup_n \underbrace{(A \cap A_n)}_{\in \mathcal{A}}$, and since μ is countably sub-additive on \mathcal{A} , we have

$$\mu(A) = \mu\left(\bigcup_n (A \cap A_n)\right) \leq \sum_n \underbrace{\mu(A \cap A_n)}_{\subseteq A_n} \leq \sum_n \mu(A_n)$$

so since the (A_n) were arbitrary, by taking infima, we have $\mu(A) \leq \mu^*(A)$.

Step 3: $\mathcal{M} \supseteq \mathcal{A}$. Let $A \in \mathcal{A}$, then $B \subseteq (B \cap A) \cup (B \cap A^c) \cup \emptyset \dots \cup \emptyset = \bigcup_n B_n$ so by (\dagger) we have

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c) + 0 \dots + 0$$

so we need to prove $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$. Again, WLOG assume $\mu^*(B) < \infty$, and so for all $\varepsilon > 0$ there exist $A_n \in \mathcal{A}$ such that $B \subseteq \bigcup_n A_n$ and

$$\mu^*(B) + \varepsilon \geq \sum_n \mu(A_n) \quad (\circ)$$

now $B \cap A \subseteq \bigcup_n \underbrace{(A_n \cap A)}_{\in \mathcal{A}}$ and $B \cap A^c \subseteq \bigcup_n \underbrace{(A_n \cap A^c)}_{A_n \setminus A \in \mathcal{A}}$. Therefore by definition of inf in μ^* and additivity of μ

$$\begin{aligned} \mu^*(B \cap A) + \mu^*(B \cap A^c) &\leq \sum_n (\mu(A \cap A_n) + \mu(A^c \cap A_n)) \\ &= \sum_n \mu(A_n) \\ &\underbrace{\leq}_{\circ} \mu^*(B) + \varepsilon \end{aligned}$$

since ε was arbitrary, $\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$, so $A \in \mathcal{M}$.

Step 4: \mathcal{M} is an algebra. Clearly $\emptyset \in \mathcal{M}$, and by the definition of \mathcal{M} its obvious that $A^c \in \mathcal{M}$ whenever $A \in \mathcal{M}$. So let $A_1, A_2 \in \mathcal{M}$

$$\mu^*(B) = \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c), \text{ since } A_1 \in \mathcal{M}$$

$$= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c), \text{ since } A_2 \in \mathcal{M}$$

Clearly $A_1 \cap A_2^c = (A_1 \cap A_2^c) \cap A_1$ and $A_1^c = (A_1 \cap A_2)^c \cap A_1^c$ so

$$\begin{aligned} \mu^*(B) &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\ &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c), \text{ since } A_1 \in \mathcal{M} \end{aligned}$$

so $A_1 \cap A_2 \in \mathcal{M}$, and \mathcal{M} is an algebra.

Step 5: Let $A = \bigcup_n A_n$, $A_n \in \mathcal{M}$, WLOG A_n disjoint (disjointification). Want $A \in \mathcal{M}$ and $\mu^*(A) = \sum_n \mu^*(A_n)$. By (\dagger) we clearly have

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c) + 0 \dots + 0$$

and

$$\mu^*(A) \leq \sum_n \mu^*(A_n)$$

so we only need two converse inequalities. Similar to before

$$\begin{aligned}
 \mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\
 &= \mu^*(B \cap A_1) + \mu^*(B \cap \underbrace{A_1^c \cap A_2}_{=A_2 \text{ as disjoint}}) + \mu^*(B \cap A_1^c \cap A_2^c) \\
 &= \sum_{n \leq N} \mu^*(B \cap A_n) + \mu^*(B \cap A_1^c \cap \dots \cap A_N^c)
 \end{aligned}$$

since $\bigcup_{n \leq N} A_n \subseteq A$ so $\bigcap_{n \leq N} A_n^c \supseteq A^c$, taking limits

$$\mu^*(B) \geq \sum_{n=1}^{\infty} \mu^*(B \cap A_n) + \mu^*(B \cap A^c)$$

and by (†)

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

so $A \in \mathcal{M}$. Applying the previous with $B = A$, we see

$$\mu^*(A) \geq \sum_{n=1}^{\infty} \mu^*(A \cap A_n) + \mu^*(A \cap A^c) = \sum_n \mu^*(A_n)$$

□

Definition. A collection \mathcal{A} of subsets of E is called a π -system if $\emptyset \in \mathcal{A}$ and if $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$.

Definition. \mathcal{A} is called a d -system if $E \in \mathcal{A}$, and if $B_1, B_2 \in \mathcal{A}$ such that $B_1 \subseteq B_2$, then $B_2 \setminus B_1 \in \mathcal{A}$, and if $A_n \in \mathcal{A}$, $A_n \uparrow \bigcup_n A_n = A$, then $A \in \mathcal{A}$.

One shows (Example sheet) that a d -system which is also a π -system is a σ -algebra.

Lemma (Dynkin). *Let \mathcal{A} be a π -system. Then any d -system that contains \mathcal{A} also contains $\sigma(\mathcal{A})$.*

Proof. Define

$$\mathcal{D} = \bigcap_{\mathcal{D}' \supseteq \mathcal{A} \text{ a } d\text{-system}} \mathcal{D}'$$

which is again a d -system (Example sheet). We show that \mathcal{D} is a π -system, hence a σ -algebra containing \mathcal{A} . Define

$$\mathcal{D}' = \{B \in \mathcal{D} : B \cap A \in \mathcal{D} \forall A \in \mathcal{A}\}$$

which contains \mathcal{A} as \mathcal{A} is a π -system. Next we show \mathcal{D}' is a d -system. Clearly $E \cap A = A \in \mathcal{A} \subseteq \mathcal{D}$, so $E \in \mathcal{D}'$. Next let $B_1, B_2 \in \mathcal{D}'$ such that $B_1 \subseteq B_2$ then $(B_2 \setminus B_1) \cap A = \underbrace{(B_2 \cap A)}_{\in \mathcal{D}} \setminus \underbrace{(B_1 \cap A)}_{\in \mathcal{D}} \in \mathcal{D}$ and so $B_2 \setminus B_1 \in \mathcal{D}'$.

Next take $B_n \uparrow B$, $B_n \in \mathcal{D}'$ then $\underbrace{B_n \cap A}_{\in \mathcal{D}} \uparrow B \cap A \in \mathcal{D}$ so $B \in \mathcal{D}'$.

Hence \mathcal{D}' is a d -system containing \mathcal{A} , so by minimality of \mathcal{D} , $\mathcal{D} \subseteq \mathcal{D}'$. Conversely, by construction $\mathcal{D}' \subseteq \mathcal{D}$, so $\mathcal{D}' = \mathcal{D}$.

Next define

$$\mathcal{D}'' = \{B \in \mathcal{D} : B \cap A \in \mathcal{D} \forall A \in \mathcal{D}\}$$

which by the preceding step ($\mathcal{D}' = \mathcal{D}$) contains \mathcal{A} . Just as before, one shows that $\mathcal{D}'' = \mathcal{D}$ and so \mathcal{D} is a π -system (as \mathcal{D}'' is by construction). \square

Theorem (Uniqueness of extension). *Let μ_1, μ_2 be measures on (E, \mathcal{E}) such that $\mu_1(E) = \mu_2(E) < \infty$, and suppose $\mu_1 = \mu_2$ on a π -system \mathcal{A} such that $\mathcal{E} \subseteq \sigma(\mathcal{A})$. Then $\mu_1 = \mu_2$ on \mathcal{E} .*

Proof. Define

$$\mathcal{D} = \{A : \mu_1(A) = \mu_2(A)\}$$

which contains \mathcal{A} by hypothesis. We show that \mathcal{D} is a d -system, and hence by Dynkin's Lemma, contains $\sigma(\mathcal{A})$, so the theorem follows.

To see this, note first that $E \in \mathcal{D}$ by hypothesis. Next, by additivity and finiteness of μ_1, μ_2 , for $B_1 \subseteq B_2$, $B_1, B_2 \in \mathcal{D}$.

$$\mu_1(B_2 \setminus B_1) = \mu_1(B_2) - \mu_1(B_1) = \mu_2(B_2) - \mu_2(B_1) = \mu_2(B_2 \setminus B_1)$$

so $B_2 \setminus B_1 \in \mathcal{D}$. Finally take $B_n \uparrow B$, $B_n \in \mathcal{D}$. This implies $B \setminus B_n \downarrow \emptyset$ and (by Example sheet) $\mu_i(B \setminus B_n) \rightarrow \mu_i(\emptyset) = 0$ for $i = 1, 2$. This implies for $\mu_i(B) < \infty$ that $\mu_i(B_n) \rightarrow \mu_i(B)$ as $n \rightarrow \infty$ for both $i = 1, 2$. But then

$$\mu_1(B) = \lim_{n \rightarrow \infty} \mu_1(B_n) = \lim_{n \rightarrow \infty} \mu_2(B_n) = \mu_2(B)$$

and so $B \in \mathcal{D}$, and thus \mathcal{D} is a d-system. \square

Remark: the above theorem applies to finite measures μ such that $\mu(E) < \infty$. The above theorem extends (as we will see) to σ -finite measures μ for which $E = \bigcup_{n \in \mathbb{N}} E_n$ such that $\mu(E_n) < \infty$.

Borel- σ -algebras

Definition. Let E be a topological space (Hausdorff, or metric space). The σ -algebra generated by $\mathcal{A} = \{B \subseteq E : B \text{ is open}\}$ is called the *Borel- σ -algebra*, denoted by $\mathcal{B}(E)$, or just \mathcal{B} when $E = \mathbb{R}$. Elements of $\mathcal{B}(E)$ are the Borel subsets of E . A measure μ on $(E, \mathcal{B}(E))$ is called a *Borel measure on E* . A *Radon* measure μ is a Borel measure such that $\mu(K) < \infty$ for all $K \subseteq E$ compact (closed in Hausdorff spaces, hence measurable).

Construction of Lebesgue measure

We will (eventually) construct a unique Borel measure μ on \mathbb{R}^d such that

$$\mu\left(\prod_{i=1}^d [a_i, b_i]\right) = \prod_{i=1}^d |b_i - a_i|, \quad a_i < b_i, \quad i = 1, \dots, d$$

We will do $d = 1$ first.

Theorem. *There exists a unique Borel measure (called the Lebesgue measure) μ on \mathbb{R} such that*

$$\mu((a, b]) = b - a, \quad \forall a < b \quad (\dagger)$$

Proof. Consider the collection \mathcal{A} of subsets of \mathbb{R} of the form

$$A = (a_1, b_1] \cup \dots \cup (a_n, b_n]$$

(intervals pairwise disjoint) which form a ring ($\emptyset = ((a, a])$, unions and differences are clear), which generates (Example sheet) generates the same σ -algebra on the open such intervals, and open intervals with rational endpoints generate \mathcal{B} , so $\sigma(\mathcal{A}) \supseteq \mathcal{B}$.

Define a set function μ on \mathcal{A} by

$$\mu(A) = \sum_{i=1}^n (b_i - a_i)$$

μ is clearly additive, and well-defined since if $A = \bigcup_j C_j$ and $A = \bigcup_k D_k$ for distinct disjoint unions, then $C_j = \bigcup_k (C_j \cap D_k)$ and $D_k = \bigcup_j (D_k \cap C_j)$, so

$$\begin{aligned} \mu(A) &= \mu\left(\bigcup_j C_j\right) = \sum_j \mu(C_j) = \sum_j \mu\left(\bigcup_k (C_j \cap D_k)\right) \\ &= \sum_{j,k} \mu(C_j \cap D_k) = \dots = \mu\left(\bigcup_k D_k\right) = \mu(A) \end{aligned}$$

by additivity of μ . Now to prove existence of μ , we apply Caratheodory's theorem and need to check that μ is countably additive on \mathcal{A} . By the Example sheet, it suffices to show that for all $A_n \in \mathcal{A}$ such that $A_n \downarrow \emptyset$ we have $\mu(A_n) \rightarrow 0$.

Assume this is not the case, so there exists some $\varepsilon > 0$ and $B_n \in \mathcal{A}$ such that $B_n \downarrow \emptyset$ but $\mu(B_n) \geq 2\varepsilon$ for all n . We can approximate B_n from within by $C_n = \bigcup_{i=1}^{N_n} \left(a_{n_i} + \frac{\varepsilon 2^{-n}}{N_n}, b_{n_i}\right] \in \mathcal{A}$ such that $\mu(B_n \setminus C_n) = \varepsilon 2^{-n} \cdot \frac{N_n}{N_n} = \varepsilon 2^{-n}$.

Now since $B_n \downarrow$, we have $B_N = \bigcap_{n \leq N} B_n$ and

$$B_N \setminus (C_1 \cap \dots \cap C_N) = B_N \cap \left(\bigcup_{n \leq N} C_n^c\right) = \bigcup_{n \leq N} B_N \setminus C_n \subseteq \bigcup_{n \leq N} B_n \setminus C_n$$

Hence since μ is increasing

$$\mu(B_N \setminus (C_1 \cap \dots \cap C_N)) \leq \mu\left(\bigcup_{n \leq N} B_n \setminus C_n\right) \leq \sum_{n \leq N} \mu(B_n \setminus C_n) \leq \varepsilon$$

Hence the “length” of what was removed $(C_1 \cap \dots \cap C_N)$ must be at least ε , i.e

$$\mu(C_1 \cap \dots \cap C_N) \geq \varepsilon > 0$$

This means that $C_1 \cap \dots \cap C_N$ is non-empty for all N , and so is

$$K_N = \overline{C_1} \cap \dots \cap \overline{C_N}$$

($\overline{C_i}$ denotes the closure of C_i) Thus K_N is a nested sequence of non-empty closed intervals, so $\emptyset \neq \bigcap_N K_N$. But $K_N \subseteq \overline{C_N} \subseteq B_N$, so $\emptyset \neq \bigcap_N K_N \subseteq \bigcap_N B_n = \emptyset$, a contradiction. So a measure μ satisfying (\dagger) must exist.

For uniqueness, suppose μ, λ measures such that (\dagger) holds, and define $\mu_n(A) = \mu(A \cap (n, n+1])$, $\lambda(A) = \lambda(A \cap (n, n+1])$ for $n \in \mathbb{Z}$, which are finite measures such that $\mu_n(E) = 1 = \lambda_n(E)$ and $\mu_n = \lambda_n$ on the π -system \mathcal{A} . So by the uniqueness theorem, we must have $\mu_n = \lambda_n$ on \mathcal{B} , and

$$\begin{aligned} \mu(A) &= \mu\left(\bigcup_n A \cap (n, n+1]\right) = \sum_n \mu(A \cap (n, n+1]) = \sum_n \mu_n(A) \\ &= \sum_n \lambda_n(A) = \dots = \lambda(A) \end{aligned}$$

so $\lambda = \mu$. □

Remarks:

1. a set $B \in \mathcal{B}$ is called a Lebesgue null set if $\mu(B) = 0$. Can write $\{x\} = \bigcap_n (x - \frac{1}{n}, x]$ and so $\mu(\{x\}) = \lim_n \frac{1}{n} = 0$. In particular $\mu((a, b)) = \mu((a, b]) = \mu([a, b])$, and any countable set Q satisfies $\mu(Q) = \mu\left(\bigcup_{q \in Q} \{q\}\right) = \sum_{q \in Q} \mu(\{q\}) = 0$. But there exist C uncountable (and measurable) in \mathcal{B} such that $\mu(C) = 0$ [Cantor set].
2. Translation invariance of μ : let $x \in \mathbb{R}$, then $B + x = \{b + x : b \in B\}$ is in \mathcal{B} whenever $B \in \mathcal{B}$ and we can define

$$\mu_x(B) = \mu(B + x)$$

and by uniqueness in the preceding theorem

$$\mu_x((a, b]) = \mu((a + x, b + x]) = (b + x) - (a + x) = b - a$$

so $\mu_x = \mu$.

3. Lebesgue-measurable sets: in the extension theorem, μ was assigned on the class \mathcal{M} , which can be shown (Example sheet) to equal

$$\mathcal{M} = \{M = A \cup N : A \in \mathcal{B}, N \subseteq B \in \mathcal{B} \text{ s.t. } \mu(B) = 0\}$$

Existence of non-measurable sets

Consider $E = (0, 1]$ with addition “+” modulo 1, and Lebesgue measure μ is still translation invariant modulo 1.

Consider the subgroup $Q = E \cap \mathbb{Q}$ of E and declare $x \sim y$ if $x - y \in Q$. This gives equivalence classes $[x] = \{y \in E : x \sim y\}$ on E . Assuming the axiom of choice, we can select a representative of $[x]$, and denote by S the set of selections running over all equivalence classes. Then we can partition E into the union of its cosets,

$$E = \bigcup_{q \in Q} (S + q)$$

a disjoint union.

Assume S is a Borel set (in $\mathcal{B}(E)$), then $S + q$ is also a Borel set for all $q \in Q$, and we can write (by countable additivity and translation invariance)

$$1 = \mu(E) = \mu\left(\bigcup_{q \in Q} (S + q)\right) = \sum_{q \in Q} \mu(S + q) = \sum_{q \in Q} \mu(S)$$

which is a contradiction. So $S \notin \mathcal{B}(E)$.

One can further show that μ cannot extend to $\mathcal{P}(E)$,

Theorem (Banach, Kuretwski). *Assuming the continuum hypothesis, there exists no measure on $([0, 1])$ such that $\mu((0, 1]) = 1$ and $\mu(\{x\}) = 0$ for all $x \in (0, 1]$.*

Proof. Not given [see Dudley, 2002]. □

Probability Spaces

If (E, \mathcal{E}, μ) (a measure space) is such that $\mu(E) = 1$, we often call it a *probability space* and write $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of outcomes/the sample space; \mathcal{F} is the set of events and \mathbb{P} is the probability measure.

The axioms of probability theory (Kolmogorov, 1933) are

1. $\mathbb{P}(\Omega) = 1$
2. $0 \leq \mathbb{P}(E) \leq 1, \forall E \in \mathcal{F}$
3. If $(A_n : n \in \mathbb{N})$ are disjoint, $A_n \in \mathcal{F}$, then $\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)$ [so \mathbb{P} is a measure on a σ -algebra]

We further say that $(A_i : i \in I)$ are *independent* if for all $J \subseteq I$ finite, we have

$$\mathbb{P} \left(\bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbb{P}(A_j)$$

We further say σ -algebras $(\mathcal{A}_i : i \in I)$ are *independent* if for any $A_j \in \mathcal{A}_j$, $j \in J$, $J \subseteq I$ finite, the A_j 's are independent.

Proposition. Let $\mathcal{A}_1, \mathcal{A}_2$ be π -systems of sets in \mathcal{F} , and suppose $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$ for all $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$. Then the σ -algebras $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2)$ are independent.

Proof. Exercise. □

The Borel-Cantelli Lemmas

For a sequence $(A_n : n \in \mathbb{N})$, $A_n \in \mathcal{F}$, define

$$\limsup_n A_n = \bigcap_n \bigcup_{m \geq n} A_m = \{A_n \text{ infinitely often "i.o."}\}$$

$$\liminf_n A_n = \bigcup_n \bigcap_{m \geq n} A_m = \{A_n \text{ eventually}\}$$

Lemma (1st Borel-Cantelli Lemma). *If $A_n \in \mathcal{F}$ are such that $\sum_n \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(A_n \text{ i.o.}) = 0$*

Proof.

$$\mathbb{P} \left(\bigcap_n \bigcup_{m \geq n} A_m \right) \leq \mathbb{P} \left(\bigcup_{m \geq n} A_m \right) \leq \sum_{m \geq n} \mathbb{P}(A_m) \rightarrow 0$$

□

Remark: the proof actually works for any measure μ .

Lemma (2nd Borel-Cantelli Lemma). *Suppose $A_n \in \mathcal{F}$ are independent and $\sum_n \mathbb{P}(A_n) = \infty$. Then $\mathbb{P}(A_n \text{ i.o.}) = 1$.*

Proof. By independence, for any $N \geq n$ and using $1 - a \leq e^{-a}$,

$$\mathbb{P} \left(\bigcap_{m=n}^N A_m^c \right) = \prod_{m=n}^N (1 - \mathbb{P}(A_m)) \leq \exp \left(- \sum_{m=n}^N \mathbb{P}(A_m) \right) \rightarrow 0 \text{ as } N \rightarrow \infty$$

Since $\bigcap_{m=n}^N A_m^c \downarrow \bigcap_{m \geq n} A_m^c$, by countable additivity we have

$$\mathbb{P} \left(\bigcap_{m \geq n} A_m^c \right) = 0$$

But then

$$\begin{aligned}\mathbb{P}(A_n \text{ i.o.}) &= \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} A_m\right) = 1 - \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} A_m^c\right) \\ &\geq 1 - \sum_n \underbrace{\mathbb{P}\left(\bigcap_{m \geq n} A_m^c\right)}_{=0} = 1\end{aligned}$$

□

2 Measurable functions

Let (E, \mathcal{E}) , (G, \mathcal{G}) be measurable spaces and let $f : E \rightarrow G$. We say that f is \mathcal{E} - \mathcal{G} -measurable if $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{G}$. If $G = \mathbb{R}$ with $\mathcal{G} = \mathcal{B}(\mathbb{R})$, we just say $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ is *measurable*.

Moreover, if E is a topological space and $\mathcal{E} = \mathcal{B}(E)$, we say f is *Borel measurable*.

Preimages preserve set operations: $f^{-1}(\bigcup_i A_i) = \bigcup_i f^{-1}(A_i)$ and $f^{-1}(G \setminus A) = E \setminus f^{-1}(A)$, which implies that $\{f^{-1}(A) : A \in \mathcal{G}\}$ is a σ -algebra over E , and likewise $\{A : f^{-1}(A) \in \mathcal{E}\}$ is also a σ -algebra over G .

This implies that if \mathcal{A} is a collection of subsets of G generating \mathcal{G} and such that $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{A}$, then $\{A : f^{-1}(A) \in \mathcal{E}\}$ is a σ -algebra containing \mathcal{A} , and hence \mathcal{G} . In particular, it suffices to check $f^{-1}(A) \in \mathcal{E}$, $\forall A \in \mathcal{A}$ to conclude that f is measurable.

If f takes real values, then

$$\mathcal{A} = \{(-\infty, y] : y \in \mathbb{R}\}$$

generates $\mathcal{B}(\mathbb{R})$ (Example sheet), and so f will be measurable whenever $f^{-1}((-\infty, y]) = \{x \in E : f(x) \leq y\} \in \mathcal{E}$ for all $y \in \mathbb{R}$. Moreover, if E is a topological space with $\mathcal{E} = \mathcal{B}(E)$, then if $f : E \rightarrow \mathbb{R}$ is continuous, it is Borel measurable.

The indicator function

$$1_A(x) = \begin{cases} 1 & \text{when } x \in A \\ 0 & \text{when } x \notin A \end{cases}$$

is measurable if and only if $A \in \mathcal{E}$.

One shows that compositions of measurable maps are measurable, and so are $f_1 + f_2$, $f_1 \cdot f_2$, $\inf_n f_n$, $\liminf_n f_n$, $\limsup_n f_n$ whenever the f_n are.

Moreover, given a collection of maps $\{f_i : E \rightarrow (G, \mathcal{G}), i \in I\}$ we can make them all measurable for

$$\sigma(f_i^{-1}(A) : A \in \mathcal{G}, i \in I)$$

Theorem (Monotone class theorem). *Let \mathcal{A} be a π -system generating the σ -algebra \mathcal{E} over E . Let further \mathcal{V} be a vector space of bounded maps from E to \mathbb{R} such that*

1. $1_E \in \mathcal{V}$, $1_A \in \mathcal{V}$, $\forall A \in \mathcal{A}$.
2. If f is bounded and $f_n \in \mathcal{V}$ is such that $0 \leq f_n \uparrow f$ pointwise on E , then $f \in \mathcal{V}$.

Then \mathcal{V} contains all bounded measurable $f : E \rightarrow \mathbb{R}$.

Proof. Define $\mathcal{D} = \{A \in \mathcal{E} : 1_A \in \mathcal{V}\}$. By hypothesis, \mathcal{D} contains the π -system \mathcal{A} and we now show it is also a d -system, so by Dynkin's lemma, $\mathcal{E} = \mathcal{D}$. Indeed, $E \in \mathcal{D}$ since $1_E \in \mathcal{V}$ by hypothesis. Also if $A \subseteq B$, $A, B \in \mathcal{D}$, then $1_{B \setminus A} = 1_B - 1_A \in \mathcal{V}$ as \mathcal{V} is a vector space. Finally, if $A_n \in \mathcal{D}$ and $A_n \uparrow A$, then $1_{A_n} \uparrow 1_A$ pointwise and so $1_A \in \mathcal{V}$ by hypothesis, so $A \in \mathcal{D}$. In particular $A \in \mathcal{V}$ for all $A \in \mathcal{E}$.

Let now $f : E \rightarrow \mathbb{R}$ be bounded, non-negative and measurable. Define

$$f_n = \sum_{j=0}^{n2^n} \frac{j}{2^n} 1_{A_{n_j}}$$

where $A_{n_j} = \{x \in E : \frac{j}{2^n} < f(x) \leq \frac{j+1}{2^n}\} = f^{-1}((\frac{j}{2^n}, \frac{j+1}{2^n}]) \in \mathcal{E}$ for $j = 0, \dots, n2^n - 1$, and $A_{n_{n2^n}} = \{x \in E : f(x) > n\} = f^{-1}((n, \infty)) \in \mathcal{E}$.

Clearly since f is bounded, for $n > \|f\|_\infty$, we see

$$f_n \leq f \leq f_n + 2^{-n}$$

so $|f_n - f| \leq 2^{-n} \rightarrow 0$. So by hypothesis $f \in \mathcal{V}$. For general f bounded and measurable, we can decompose $f = f^+ - f^-$ where $f^\pm \geq 0$, and repeat the argument above. \square

Image Measures

If $f : (E, \mathcal{E}) \rightarrow (G, \mathcal{G})$ is \mathcal{E} - \mathcal{G} measurable, and μ is a measure on \mathcal{E} , then the image measure $\nu = \mu \circ f^{-1}$ is obtained from

$$\nu(A) = \mu(f^{-1}(A)), \quad \forall A \in \mathcal{G}$$

which is indeed a measure on \mathcal{G} (Example sheet).

Lemma. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a right-continuous, monotone increasing function, and set $g(\pm\infty) = \lim_{z \rightarrow \pm\infty} g(z)$. On $I = (g(-\infty), g(\infty))$ define

$$f(x) = \inf\{y \in \mathbb{R} : x \leq g(y)\}, \quad x \in I$$

Then f is monotone increasing, left-continuous and

$$f(y) \leq y \iff x \leq g(y) \quad \forall x, y$$

Proof. Define $J_x = \{y \in \mathbb{R} : x \leq g(y)\}$. Since $x > g(-\infty)$, J_x is non-empty and bounded below, so $f(x) \in \mathbb{R}$. Now if $y \in J_x$ then $y' \geq y$ implies $y' \in J_x$ as well since $g \uparrow$. Moreover if $y_n \downarrow y$, $y_n \in J_x$, then we can take limits in $x \leq g(y_n)$ to see $x \leq \lim_n g(y_n) = g(y)$ as g is right-continuous, so $y \in J_x$. We conclude that $J_x = [f(x), \infty)$, which shows the equivalence.

Moreover, if $x \leq x'$, then $J_x \supseteq J_{x'}$ since $g \uparrow$. So by properties of the infimum $f(x) \leq f(x')$. Likewise if $x_n \uparrow x$, then $J_x = \bigcap_n J_{x_n}$ so $f(x_n) \rightarrow f(x)$ as $x_n \rightarrow x$. \square

We call f the *generalised inverse of g* .

Theorem. Let g be as in the above lemma. Then there exists a unique Radon measure μ_g on \mathbb{R} such that $\mu_g((a, b]) = g(b) - g(a)$ for all $a < b$. Every Radon measure on \mathbb{R} can be obtained in this way.

Proof. For f as defined in the previous lemma, note that for all $z \in \mathbb{R}$

$$f^{-1}((-\infty, z]) = \{x : f(x) \leq z\} = \{x : x \leq g(y)\} = (g(-\infty), g(z)] \in \mathcal{B}(I)$$

Where the 2nd equality follows again from the lemma. So f is $\mathcal{B}\text{-}\mathcal{B}(I)$ measurable, and the image measure $\mu \circ f^{-1} = \mu_g$, where μ is the Lebesgue measure on I , exists.

Then for $-\infty < a < b < \infty$ we have

$$\mu_g((a, b]) = \mu(f^{-1}((a, b])) = \mu(x \in I : a < f(x) \leq b) = \mu((g(a), g(b)]) = g(b) - g(a)$$

Which uniquely determines μ_g by the same arguments as for the Lebesgue measure on \mathbb{R} . (Since g maps into \mathbb{R} , μ_g is a Radon measure).

Conversely, let ν be any Radon measure on \mathbb{R} , define

$$g(y) = \begin{cases} \nu((0, y]) & y \geq 0 \\ -\nu((y, 0]) & y < 0 \end{cases}$$

Which is clearly increasing in y (since ν is increasing). If $y_n \downarrow y$, then $(0, y_n] \downarrow (0, y]$ so $g(y_n) \rightarrow g(y)$ since ν is countably additive, so g is right-continuous. Finally (assuming $a < 0 < b$, the other cases are similar),

$$\nu((a, b]) = \nu((a, 0]) + \nu((0, b]) = -g(a) + g(b) = g(b) - g(a)$$

And by uniqueness as before, the result follows. \square

Remark: The μ_g are called Lebesgue-Stieltjes measures, with Stieltjes distribution g .

For example, the Dirac measure δ_x at $x \in \mathbb{R}$, defined by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Which has Stieltjes distribution $g = 1_{[x, \infty)}$.

Random Variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and (E, \mathcal{E}) a measurable space.

Definition. An E -valued random variable X is any \mathcal{F} - \mathcal{E} measurable map

$$X : \Omega \rightarrow E$$

When $E = \mathbb{R}, \mathbb{R}^d$ (with Borel σ -algebras) we call X a *random variable*, or random vector. The *law* or *distribution* μ_X of a random variable is given by $\mu_X = \mathbb{P} \circ X^{-1}$ (the image measure) with, for $E = \mathbb{R}$ distribution function

$$F_X(z) = \mu_X((-\infty, z]) = \mathbb{P}(X^{-1}((-\infty, z])) = \mathbb{P}(\omega \in \Omega : X(\omega) \leq z) = \mathbb{P}(X \leq z)$$

which uniquely determines μ_X .

Using properties of measures one shows that any distribution function satisfies

1. $F_X \uparrow$
2. F_X is right-continuous
3. $\lim_{z \rightarrow -\infty} F_X(z) = \mu_X(\emptyset) = 0$ and $\lim_{z \rightarrow \infty} F_X(z) = \mu_X(\mathbb{R}) = \mathbb{P}(\Omega) = 1$

Given any distribution function F_X satisfying 1,2 & 3, we can on $(\Omega, \mathcal{F}, \mathbb{P}) = ((0, 1), \mathcal{B}(0, 1), \mu)$, where μ is the Lebesgue measure obtain a random variable $X : \Omega \rightarrow \mathbb{R}$ by

$$X(\omega) = \inf\{x : \omega \leq F_X(x)\}$$

with distribution function F_X .

Definition. A countable collection $(X_i : (\Omega, \mathcal{F}, \mathbb{P} \rightarrow (E, \mathcal{E}))$ of random variables is said to be *independent* whenever the σ -algebras $\sigma(X_i^{-1}(A) : A \in \mathcal{E})$ are independent. For $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ one shows (Example sheet) that this is equivalent (for $I = \{1, \dots, n\}$) to

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i), \quad \forall x_i \in \mathbb{R}$$

We now construct on $(\Omega, \mathcal{F}, \mathbb{P}) = ((0, 1), \mathcal{B}, \mu|_{(0,1)})$ with $\mu|_{(0,1)}$ the Lebesgue measure on $(0, 1)$ an infinite sequence of independent random variables with prescribed distribution functions F_n .

Any $\omega \in (0, 1)$ has a binary representation $(\omega_i) \in \{0, 1\}^{\mathbb{N}}$, where $\omega = \sum_{i=1}^{\infty} \omega_i 2^{-i}$, which is unique if we exclude sequences which terminate with infinitely many 0's (so rationals end in a sequence of 1's). Then we can define $R_n(\omega) = \omega_n$ ("Radenmacher functions"), which are of the form

$$\begin{aligned} R_1(\omega) &= 1_{(1/2, 1)} \\ R_2(\omega) &= 1_{(1/4, 1/2]} + 1_{(3/4, 1)} \\ R_3(\omega) &= 1_{(1/8, 1/4]} + 1_{(3/8, 1/2]} + 1_{(5/8, 3/4]} + 1_{(7/8, 1)} \end{aligned}$$

So the R_n are random variables such that $\mathbb{P}(R_n = 1) = \frac{1}{2} = \mathbb{P}(R_n = 0)$, so the R_n are Bernoulli for all n . Moreover for $(x_i)_{i=1}^n \in \{0, 1\}^n$

$$\mathbb{P}(R_1 = x_1, \dots, R_n = x_n) = 2^{-n} = \underbrace{\mathbb{P}(R_1 = x_1) \dots \mathbb{P}(R_n = x_n)}_{\frac{1}{2}}$$

So the R_n are all independent. Now take a bijection $m : \mathbb{N}^2 \rightarrow \mathbb{N}$ and define $Y_{nk} = R_{m(n,k)}$ which are again independent and define

$$Y_n = \sum_k 2^{-k} Y_{nk}$$

which converge for all $\omega \in \Omega$ since $|Y_{nk}| \leq 1$ are still independent. To determine the law of Y_n we consider the π -system of intervals $(\frac{i}{2^m}, \frac{i+1}{2^m}]$, $i = 0, \dots, 2^m - 1$, $m \in \mathbb{N}$, with dyadic endpoints, which generate \mathcal{B} and

$$\begin{aligned} \mathbb{P}\left(Y_n \in \left(\frac{i}{2^m}, \frac{i+1}{2^m}\right]\right) &= \mathbb{P}\left(\frac{i}{2^m} < \sum_k 2^{-k} Y_{nk} \leq \frac{i+1}{2^m}\right) = 2^{-m} \\ &= \mu|_{(0,1)}\left(\left(\frac{i}{2^m}, \frac{i+1}{2^m}\right]\right) \end{aligned}$$

so the law $\mu_{Y_n} = \mu|_{(0,1)}$ by the uniqueness theorem, and so the Y_n 's are an infinite sequence of independent uniform random variables. Now if F_n are probability distribution functions (satisfy axioms 1-3 from earlier), then taking the generalised inverse $f_n = F_n^{-1}$ from the lemma, we see that the $F_n^{-1}(Y_n)$ are independent and have distribution function F_n .

Convergence of measurable functions

Definition. We say that a property defining a set $A \in \mathcal{E}$ holds μ -almost everywhere if $\mu(A^c) = 0$ for a measure μ on \mathcal{E} . If $\mu = \mathbb{P}$, we say it holds \mathbb{P} -almost surely, or with probability 1, if $\mathbb{P}(A) = 1$.

If f_n, f are measurable maps on $(E, \mathcal{E}|_\mu)$ we say $f_n \rightarrow f$ μ -almost always if

$$\mu(x \in E : f_n(x) \not\rightarrow f(x) \text{ as } n \rightarrow \infty) = 0$$

We say $f_n \rightarrow f$ in μ -measure if for all $\varepsilon > 0$

$$\mu(x \in E : |f_n(x) - f(x)| > \varepsilon) \rightarrow 0 \text{ as } n \rightarrow \infty$$

For random variables say $X_n \rightarrow X$ \mathbb{P} -almost surely or $X_n \rightarrow X$ in \mathbb{P} -probability respectively.

If $E = \mathbb{R}$, we say $X_n \xrightarrow{d} X$ in distribution if $\mathbb{P}(X_n \leq x) \rightarrow \mathbb{P}(X \leq x)$ for all $x \in \mathbb{R}$ such that $x \mapsto \mathbb{P}(X \leq x)$ is continuous. One shows $X_n \xrightarrow{P} X$ implies $X_n \xrightarrow{d} X$.

Theorem. Let $f_n : (E, \mathcal{E}) \rightarrow \mathbb{R}$ be measurable functions.

1. If $\mu(E) < \infty$, then whenever $f_n \rightarrow 0$ a.e (almost everywhere) we have $f_n \rightarrow 0$ in measure.
2. If $f_n \rightarrow 0$ in measure, then $f_{n_k} \rightarrow 0$ a.e along some subsequence n_k .

Proof.

1. For all $\varepsilon > 0$ we have

$$\begin{aligned} \mu(|f_n| \leq \varepsilon) &\geq \mu\left(\bigcap_{m \geq n} \underbrace{\{|f_m| \leq \varepsilon\}}_{:= A_m}\right) \\ &\uparrow \mu\left(\bigcup_n \bigcap_{m \geq n} A_m\right) \\ &= \mu(|f_n| \leq \varepsilon \text{ eventually}) \\ &\geq \mu(f_n \rightarrow 0 \text{ as } n \rightarrow \infty) \\ &= \mu(E) \end{aligned}$$

so $\liminf_n \mu(|f_n| \leq \varepsilon) \geq \mu(E)$. So we see $\limsup_n \mu(|f_n| > \varepsilon) \leq \mu(E) - \mu(E) = 0$, so $\mu(|f_n| > \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$ as desired.

2. By hypothesis, for all $\varepsilon > 0$ $\mu(|f_n| > \frac{1}{k}) < \varepsilon$ for n large enough. So choosing $\varepsilon = \frac{1}{k^2}$ we see that along some subsequence n_k we have $\mu(|f_{n_k}| > \frac{1}{k}) \leq \frac{1}{k^2}$ so

$$\sum_k \mu(|f_{n_k}| > \frac{1}{k}) < \infty$$

and by the 1st Borel-Cantelli Lemma, we have $\mu(|f_{n_k}| > \frac{1}{k} \text{ i.o.}) = 0$, so $f_{n_k} \rightarrow 0$ a.e.

□

Remarks: (1) is false if $\mu(E) = \infty$, as the example $1_{(n,\infty)}$ on $(\mathbb{R}, \mathcal{B}, \mu)$, μ Lebesgue measure shows. (2) is false without restricting to subsequences: take A_n independent such that $\mathbb{P}(A_n) = \frac{1}{n}$ then $1_{A_n} \rightarrow 0$ in \mathbb{P} -probability since $\mathbb{P}(1_{A_n} > \varepsilon) = \mathbb{P}(A_n) = \frac{1}{n} \rightarrow 0$ but $\sum_n \mathbb{P}(A_n) = \infty$, so by the 2nd Borel-Cantelli Lemma, $\mathbb{P}(1_{A_n} > \varepsilon \text{ i.o.}) = 1$, so $1_{A_n} \not\rightarrow 0$ a.s.

Example. Let $(X_n : n \in \mathbb{N})$ be independent and identically distributed (iid) exponential random variables with $\mathbb{P}(X_1 \leq x) = 1 - e^{-x}$, $x \geq 0$. Define $A_n = \{X_n \geq \alpha \log n\}$, $\alpha > 0$, s.t $\mathbb{P}(A_n) = n^{-\alpha}$ and $\sum_n \mathbb{P}(A_n) < \infty$ if and only if $\alpha > 1$. So by the Borel-Cantelli lemmas, we have

$$\mathbb{P}\left(\frac{X_n}{\log n} \geq 1 \text{ i.o.}\right) = 1$$

while

$$\mathbb{P}\left(\frac{X_n}{\log n} \geq 1 + \varepsilon \text{ i.o.}\right) = 0 \quad \forall \varepsilon > 0$$

So $\limsup_n \frac{X_n}{\log n} = 1$ almost surely.

Kolmogorov's 0-1 Law

For $(X_n : n \in \mathbb{N})$ random variables, define $\mathcal{T} = \sigma(X_{n+1}, X_{n+2}, \dots)$ and set $\mathcal{T} = \bigcap_{n \in \mathbb{N}} \mathcal{T}_n$, the “tail σ -algebra” which contains all events in \mathcal{F} which depend only on the limiting behaviour of the sequence.

Theorem. For $(X_n : n \in \mathbb{N})$ independent random variables, if $A \in \mathcal{T}$ then $\mathbb{P}(A) = 1$ or $\mathbb{P}(A) = 0$. Moreover if $Y : (\Omega, \mathcal{T}) \rightarrow (\mathbb{R}, \mathcal{B})$ is measurable, then Y is constant almost surely.

Proof. Define $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ which is a σ -algebra generated by the π -system of sets

$$A = (X_1 \leq x_1, \dots, X_n \leq x_n), \quad x_i \in \mathbb{R}$$

and note that the π -system of sets

$$B = (X_{n+1} \leq x_{n+1}, \dots, X_{n+k} \leq x_{n+k}), \quad k \in \mathbb{N}, \quad x_i \in \mathbb{R}$$

generates \mathcal{T}_n . By independence of X_n , $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, so by the theorem from earlier we see that \mathcal{T}_n and \mathcal{F}_n are independent. If we set $\mathcal{F}_\infty = \sigma(X_1, X_2, \dots)$, then $\bigcup_n \mathcal{F}_n$ is a π -system generating \mathcal{F}_∞ , and if $A \in \bigcup_n \mathcal{F}_n$, there exists \bar{n} such that $B \in \mathcal{T}_{\bar{n}}$ is independent of A , in particular A is independent of elements in $\mathcal{T} = \bigcap_{\bar{n}} \mathcal{T}_{\bar{n}}$, hence as before \mathcal{F}_∞ is independent of \mathcal{T} . But clearly $\mathcal{T} \subseteq \mathcal{F}_\infty$, so if $A \in \mathcal{T}$ it is independent to $A \in \mathcal{F}_\infty$! Now $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, so $\mathbb{P}(A) = 0$ or 1 . Finally, if Y is \mathcal{T} measurable, then $\{Y \leq y\}$ lies in \mathcal{T} for all y , hence have probability 1 or 0. Then let

$$c = \inf\{y : F_Y(y) = 1\}$$

so $Y = c$ almost surely. □

3 Integration

For $f : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ measurable or “integrable” we will define the integral with respect to μ :

$$\mu(f) = \int_E f d\mu = \int_E f(x) d\mu(x)$$

and if X is a random variable, we define its (“mathematical”) expectation as

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

To start, call $f : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ *simple* if it is of the form

$$f = \sum_{k=1}^m a_k 1_{A_k}, \quad a_k \geq 0, \quad A_k \in \mathcal{E}, \quad m \in \mathbb{N}$$

We define its μ -integral to be

$$\mu(f) = \sum_{k=1}^m a_k \mu(A_k)$$

which is well-defined (Example sheet) and it satisfies the following properties:

1. $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$ for all $\alpha, \beta \geq 0$ and f, g simple
2. If $g \leq f$ then $\mu(g) \leq \mu(f)$
3. If $f = 0$ almost everywhere $\mu(f)$

For general $f : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ non-negative, we define its μ -integral as

$$\mu(f) = \sup\{\mu(g) : g \leq f, \text{ } g \text{ simple}\}$$

which is consistent with the definition for simple functions, and takes values in $[0, \infty]$.

For $f : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ measurable (but not necessarily non-negative), we define $f^+ = \max(f, 0)$, $f^- = \max(-f, 0)$, so that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. We say that f is μ -integrable if $\mu(|f|) < \infty$. In this case we define

$$\mu(f) = \mu(f^+) - \mu(f^-)$$

which is well-defined (i.e not $\infty - \infty$).

Theorem (Monotone Convergence Theorem). *Let $f_n, f : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ be measurable and non-negative such that $0 \leq f_n \uparrow f$ (i.e $f_n(x) \leq f_{n+1}(x) \leq f(x)$ and $f_n(x) \rightarrow f(x)$ for all $x \in E$). Then $\mu(f_n) \rightarrow \mu(f)$ as $n \rightarrow \infty$.*

Remark: if we take the approximating sequence $\tilde{f}_n (= \min(2^{-n} \lfloor 2^n f \rfloor, n))$ then $0 \leq \tilde{f} \uparrow f$ so $\mu(f) = \lim_n \mu(\tilde{f}_n)$.

Proof. Recall $\mu(f) = \sup\{\mu(g) : g \leq f, g \text{ simple}\}$. Since $0 \leq f_n \uparrow$ we have $\mu(f_n) \uparrow \sup_n \mu(f_n) = M$. But then since $f_n \leq f$ we must have $\mu(f_n) \leq \mu(f)$ so taking suprema $M \leq \mu(f)$, and if $M < \infty$ we have $\lim_n \mu(f_n) \leq \mu(f)$.

We will now show $\mu(g) \leq M$ for all simple functions g such that $g \leq f$ so that taking suprema $\mu(f) = \sup_g \mu(g) \leq M$ so $\mu(f) = \lim_n \mu(f_n)$ follows.

We define $g_n = \min(\bar{f}_n, g) = \bar{f}_n \wedge g$, where \bar{f}_n is the approximation of f_n by simple functions from the monotone class theorem, $[\bar{f}_n]_n = \bar{f}_n = \min(2^{-n} \lfloor 2^n f_n \rfloor, n)$. Now since $f_n \uparrow f$ we must have $\bar{f}_n \uparrow f$ too, and so $g_n \uparrow \min(f, g) = g$, and since $\bar{f}_n \leq f_n$ we also have $g_n \leq f_n$ for all n .

Now let g be an arbitrary simple function, of the form

$$g = \sum_{k=1}^m a_k 1_{A_k}$$

with $m \in \mathbb{N}$, $a_k \geq 0$ and $A_k \in \mathcal{E}$ disjoint (wlog). We define for $\varepsilon > 0$ arbitrary

$$A_k(n) = \{x \in A_k : g_n(x) \geq (1 - \varepsilon)a_k\}$$

Since $g = a_k$ on A_k and since $g_n \uparrow g$, we have $A_k(n) \uparrow A_k$ for all k . Also since μ is a measure, we must have $\mu(A_k(n)) \uparrow \mu(A_k)$. We have $g_n 1_{A_k} \geq g_n 1_{A_k(n)} \geq (1 - \varepsilon)a_k 1_{A_k(n)}$ on E . Moreover

$$g_n = \sum_{k=1}^m g_n 1_{A_k}$$

since the A_k 's are disjoint and support g_n (if $1_{A_n} = 0$ for all n , then $g = 0$ and $f_n = 0$). Now

$$\mu(g_n) = \sum_{k=1}^m \mu(g_n 1_{A_k}) \geq (1 - \varepsilon) \sum_{k=1}^m a_k \mu(A_k(n)) \uparrow (1 - \varepsilon) \sum_{k=1}^m a_k \mu(A_k) = (1 - \varepsilon)\mu(g)$$

So $\mu(g) \leq \frac{1}{1 - \varepsilon} \limsup_n \mu(g_n) \leq \frac{1}{1 - \varepsilon} \limsup_n \mu(f_n) \leq \frac{M}{1 - \varepsilon}$. Since ε was arbitrary we have $\mu(g) \leq M$ as required. \square

Remarks: we have shown $\mu(f) = \mu(\lim_n f_n) = \lim_n \mu(f_n)$, so we can interchange $\int(\cdot)d\mu$ and the limit. If $g_n \geq 0$, then $\mu(\sum_n g_n) = \sum_n \mu(g_n)$. Moreover it suffices to require $f_n \uparrow f$ almost everywhere and the $f_n \geq 0$ hypothesis is not necessary as long as f_1 is integrable (then just subtract f_1 from all terms).

Theorem. Let $f, g : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ be measurable and non-negative. Then

1. $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$ for all $\alpha, \beta \geq 0$
2. If $g \leq f$ then $\mu(g) \leq \mu(f)$

3. $f = 0$ almost everywhere if and only if $\mu(f) = 0$.

Proof. If \tilde{f}_n, \tilde{g}_n are the approximations of f, g from the monotone class theorem, then $\alpha\tilde{f}_n \uparrow \alpha f$, $\beta\tilde{g}_n \uparrow \beta g$, $\alpha\tilde{f}_n + \beta\tilde{g}_n \uparrow \alpha f + \beta g$. And from earlier

$$\mu(\alpha\tilde{f}_n + \beta\tilde{g}_n) = \alpha\mu(\tilde{f}_n) + \beta\mu(\tilde{g}_n)$$

So taking limits the monotone convergence theorem implies

$$\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$$

(2) follows in a similar way. Now we show (3): if $f = 0$ almost everywhere, then $0 \leq \tilde{f}_n \leq f = 0$ a.e, so $\tilde{f}_n = 0$ a.e for all n , so $\mu(\tilde{f}_n) = 0$, so $\mu(\tilde{f}_n) \uparrow \mu(f) = 0$. Conversely if $\mu(f) = 0$ then $0 \leq \mu(\tilde{f}) \uparrow \mu(f) = 0$ so $\mu(\tilde{f}_n) = 0$ for all n , so $\tilde{f}_n = 0$ a.e. Since $0 \leq \tilde{f}_n \uparrow f$ we have that $f = 0$ a.e. \square

Remark: functions such as $1_{\mathbb{Q}}$ have $\mu(1_{\mathbb{Q}}) = 0$, and are ‘identified’ with 0.

Theorem. Let $f, g : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ be integrable. Then

1. $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$ for all $\alpha, \beta \in \mathbb{R}$
2. $g \leq f$ implies $\mu(g) \leq \mu(f)$
3. If $f = 0$ almost everywhere then $\mu(f) = 0$

Proof. Clearly if f is integrable, so is αf , and $\mu(-f) = -\mu(f)$. And for $\alpha \geq 0$, $\mu(\alpha f) = \mu((\alpha f)^+) - \mu((\alpha f)^-) = \alpha\mu(f^+) - \alpha\mu(f^-) = \alpha\mu(f)$. So we can restrict to $\alpha = \beta = 1$.

Define $h = f + g = h^+ - h^- = f^+ - f^- + g^+ - g^-$. This is the same as $h^+ + f^- + g^- = h^- + f^+ + g^+$, and all of these functions are non-negative. Hence by the previous theorem

$$\mu(h^+) + \mu(f^-) + \mu(g^-) = \mu(h^-) + \mu(f^+) + \mu(g^+)$$

so $\mu(h) = \mu(f) + \mu(g)$ follows.

Now we show (2). Clearly $0 \leq f - g$ so $\mu(0) \leq \mu(f - g)$ by the previous theorem, and $\mu(f - g) = \mu(f) - \mu(g)$ by (1) of this theorem.

Finally we show (3): if $f = 0$ almost everywhere, $f^+ = f^- = 0$ almost everywhere, so $\mu(f) = \mu(f^+) - \mu(f^-) = 0 - 0$. \square

Lemma (Fatou). *Let $f_n, f : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ be measurable and non-negative. Then $\mu(\liminf_n f_n) \leq \liminf_n \mu(f_n)$.*

Remark: recall that for $x_n \in \mathbb{R}$

$$\liminf_n x_n = \sup_n \inf_{m \geq n} x_m$$

$$\limsup_n x_n = \inf_n \sup_{m \geq n} x_m$$

In particular, if $\limsup_n x_n = \liminf_n x_n$ then $\lim_n x_n = \liminf_n x_n$. Therefore if $f = \lim_n f_n$ exists in Fatou's lemma, we have $\mu(f) \leq \liminf_n \mu(f_n)$.

Proof. We have $\inf_{m \geq n} f_m \leq f_k$ for all $k \geq n$, and integrating this implies $\mu(\inf_{m \geq n} f_m) \leq \mu(f_k)$ for all $k \geq n$. So

$$\mu(\inf_{m \geq n} f_m) \leq \mu(f_k)$$

$$\mu(\inf_{m \geq n} f_m) \leq \inf_{k \geq n} \mu(f_k) \leq \sup_n \inf_{k \geq n} \mu(f_k) = \liminf_n \mu(f_n)$$

Also, $0 \leq \inf_{m \geq n} f_m \uparrow \sup_n \inf_{m \geq n} f_m$ so by the monotone convergence theorem

$$\mu(\liminf_n f_n) = \lim_n \mu(\inf_{m \geq n} f_m) \leq \liminf_n \mu(f_n)$$

□

Theorem (Dominated convergence theorem). *Let $f_n, f : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ be measurable such that $|f_n| \leq g$ almost everywhere on E and g is μ -integrable ($\mu(g) < \infty$). Suppose $f_n \rightarrow f$ pointwise (or almost everywhere) on E . Then f_n and f are integrable and $\mu(f_n) \rightarrow \mu(f)$ as $n \rightarrow \infty$.*

Proof. Clearly $\mu(|f_n|) \leq \mu(g) < \infty$ so the f_n are integrable and taking limits in $|f_n| \leq g$ we have $|f| \leq g$, so $\mu(|f|) < \infty$.

Next

$$0 \leq g \pm f_n \xrightarrow{\text{ptws on } E} g \pm f$$

By Fatou's lemma

$$\mu(g) + \mu(f) = \mu(g + f) = \mu(\liminf_n (g + f_n)) \leq \liminf_n (\mu(g) + \mu(f_n)) = \mu(g) + \liminf_n \mu(f_n)$$

So $\mu(f) \leq \liminf_n \mu(f_n)$. Likewise

$$\mu(g) - \mu(f) = \mu(\liminf_n (g - f_n)) \leq \mu(g) - \limsup_n \mu(f_n)$$

So $\limsup_n \mu(f_n) \leq \mu(f)$. Therefore $\limsup_n \mu(f_n) = \liminf_n \mu(f_n) = \lim_n \mu(f_n) = \mu(f)$. □

Example. On $E = [0, 1]$ with the Lebesgue measure, suppose $f_n \rightarrow f$ pointwise and $\sup_n \|f_n\|_\infty \leq g < \infty$. Then since $\mu(g) \leq g$ the dominated convergence theorem implies $\mu(f_n) \rightarrow \mu(f)$ as $n \rightarrow \infty$ (no uniform convergence of $f_n \rightarrow f$ required).

Remark: the proof of the Fundamental Theorem of Calculus (FTC) requires only $\int_x^{x+h} dt = h$. Therefore for any continuous $f : [0, 1] \rightarrow \mathbb{R}$

$$\underbrace{\int_0^x f(t) dt}_{\text{Riemann-integral}} = F(x) = \underbrace{\int_0^x f(t) d\mu(t)}_{\text{Lebesgue-integral}}, \quad x \in [0, 1]$$

So these integrals coincide for continuous maps.

One shows that all Riemann-integrable functions are μ^* -measurable (μ is Lebesgue measure) but that there exist Riemann-integrable functions that are not Borel measurable.

A bounded μ^* -measurable function is Riemann-integrable if and only if

$$\mu(\{x \in [0, 1] : f \text{ is discontinuous at } x\}) = 0$$

All standard formulae for the Riemann-integral (substitution, integration, by parts etc) extend to all bounded measurable functions by the monotone class theorem (see Example sheet).

Theorem. Let $U \subseteq \mathbb{R}$ be open, (E, \mathcal{E}, μ) a measure space, and $f : U \times E \rightarrow \mathbb{R}$ such that

- $x \mapsto f(t, x)$ for all $t \in U$ is measurable
- $t \mapsto f(t, x)$ is differentiable for all $x \in E$, with $\left| \frac{\partial f(t, x)}{\partial t} \right| \leq g(x)$ for all $t \in U$ where g is μ -integrable.

Then if

$$F(t) = \int_E f(t, x) d\mu(x)$$

we have

$$F'(t) = \int_E \frac{\partial f}{\partial t}(t, x) d\mu(x)$$

Proof. By the MVT

$$|g_h(x)| := \left| \frac{f(t+h, x) - f(t, x)}{h} - \frac{\partial f}{\partial t}(t, x) \right| = \left| \frac{\partial f(\tilde{t}, x)}{\partial t} - \frac{\partial f(t, x)}{\partial t} \right|$$

For some $\tilde{t} \in U$. SO $|g_h(x)| \leq 2g(x)$ which is μ -integrable. By differentiability, we have $g_h \rightarrow 0$ as $h \rightarrow 0$, so applying the dominated convergence theorem, $\mu(g_h) \rightarrow \mu(0) = 0$, or by linearity of μ

$$\mu(g_h) = \frac{\int_E (f(t+h, x) - f(t, x)) d\mu(x)}{h} - \int_E \frac{\partial f}{\partial t}(t, x) d\mu(x)$$

$$= \frac{F(t+h) - F(t)}{h} - F'(t) \rightarrow 0 \text{ as } h \rightarrow 0$$

□

Integrals with respect to image measures

For $f : (E, \mathcal{E}, \mu) \rightarrow (G, \mathcal{G})$ measurable, $g : G \rightarrow \mathbb{R}$ measurable and non-negative, we have

$$\mu \circ f^{-1}(g) = \int_G g d\mu \circ f^{-1} = \int_E g(f(x)) d\mu(x) = \mu(g \circ f)$$

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and for a G -valued random variable X ,

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

Measures with densities

If $f : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ is measurable and non-negative, we can define $\nu_f(A) = \mu(f1_A)$ for any $A \in \mathcal{E}$, which is again a measure (by the monotone convergence theorem), and if $g : (E, \mathcal{E}) \rightarrow \mathbb{R}$ is measurable, then $\nu_f(g) = \int_E g(x)f(x)d\mu(x) = \int_E g d\nu_f$. We call f the density of ν_f with respect to μ .

Product measures

Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be finite measure spaces. On $E = E_1 \times E_2$, we consider the π -system of ‘rectangles’ $\mathcal{A} = \{A_1 \times A_2 : A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$, which generates the σ -algebra $\sigma(\mathcal{A}) \equiv \mathcal{E}_1 \otimes \mathcal{E}_2 = \mathcal{E}$.

If E_1, E_2 are topological spaces with a countable base, then $\mathcal{B}(E_1 \times E_2)$ for the product topology on $E_1 \times E_2$ coincides with $\mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$ (see Dudley).

Lemma. *Let $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ be measurable. Then for all $x_1 \in E_1$ fixed the map $x_2 \mapsto f(x_1, x_2)$ is \mathcal{E}_2 -measurable.*

Proof. Define a vector space

$$\mathcal{V} = \{f : (E, \mathcal{E}) \rightarrow \mathbb{R} \text{ bounded and measurable, s.t Lemma holds}\}$$

This is indeed a vector space, and contains $1_E, 1_A$ for all $A \in \mathcal{A}$, since $1_A = 1_{A_1}(x_1)1_{A_2}(x_2)$ is \mathcal{E}_2 measurable as $A_2 \in \mathcal{E}_2$. Next let $0 \leq f_n \uparrow f$, $f_n \in \mathcal{V}$, then $f(x_1, \cdot) = \lim_n f_n(x_1, \cdot)$ hence is \mathcal{E}_2 -measurable as the limit of a sequence of measurable functions, so by the monotone class theorem, \mathcal{V} contains all bounded measurable functions. This extends to all f (not necessarily bounded) by taking $\min(\max(-n, f), n) \in \mathcal{V}$, which converges to f . \square

Lemma. *Let $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ be measurable, such that either*

1. *f is bounded or;*
2. *$f \geq 0$*

Then $x_1 \mapsto \int_{E_2} f(x_1, x_2) d\mu_2(x_2)$ is \mathcal{E}_1 measurable, and is (in the case of 1) bounded on E_1 , (in the case of 2) ≥ 0 , respectively.

Remarks: in 2, the mapping may evaluate to ∞ , but $\{x_1 \in E_1 : \int_{E_2} f(x_1, x_2) d\mu(x_2) = \infty\} \in \mathcal{E}_1$

Proof. Define a vector space

$$\mathcal{V} = \{f : (E, \mathcal{E}) \rightarrow \mathbb{R} \text{ bounded and measurable, s.t Lemma holds}\}$$

Which is indeed a vector space, and contains 1_E since $1_{E_1} \mu(E_2) \geq 0$ is bounded and also $1_A = 1_{A_1}(x_1) 1_{A_2}(x_2)$ since $1_{A_1}(x_1) \mu_2(A_2)$ is \mathcal{E}_1 -measurable, non-negative and bounded since $0 \leq \mu_2(E_2) < \infty$.

Now let $0 \leq f_n \uparrow f$ be a sequence in \mathcal{V} . Then by the monotone convergence theorem,

$$\int_{E_2} \lim_n f_n(x_1, x_2) d\mu_2(x_2) = \lim_n \int_{E_2} f_n(x_1, x_2) d\mu_2(x_2)$$

which is \mathcal{E}_1 -measurable as the limit of \mathcal{E}_1 -measurable functions. Also (in the case of 1) it is bounded by $\mu_2(E_2) \|f\|_\infty$ and non-negative, so $f \in \mathcal{V}$, so by the monotone class theorem, \mathcal{V} contains all bounded measurable functions. In the case of 2, we approximate f by $\min(f, n) \in \mathcal{V}$. \square

Theorem (Product measure). *Let $\mu_1(E_1), \mu_2(E_2) < \infty$. Then there exists a unique measure μ on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ such that $\mu(A_1 \times A_2) = \mu_1(A_1) \mu_2(A_2)$ for all $A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2$.*

Proof. By the uniqueness theorem and since \mathcal{A} generates $\mathcal{E}_1 \otimes \mathcal{E}_2$, there can only be one such measure. Define

$$\mu(A) = \int_{E_1} \left(\int_{E_2} 1_A(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1)$$

, so $\mu(A_1 \times A_2) = \int_{E_1} 1_{A_1}(x_1) \mu_2(A_2) d\mu_1(x_1) = \mu_1(A_1) \mu_2(A_2)$, and $\mu(\emptyset) = 0$, so to prove the theorem we need to show μ is countably additive (and thus a measure). Let $A_n \in \mathcal{E}_1 \otimes \mathcal{E}_2$ be disjoint, so $1_{\bigcup_n A_n} = \sum_n 1_{A_n} = \lim_{N \rightarrow \infty} \sum_{n \leq N} 1_{A_n}$. Thus

$$\mu\left(\bigcup_n A_n\right) = \int_{E_1} \left(\int_{E_2} \lim_{N \rightarrow \infty} \sum_{n \leq N} 1_{A_n}(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1)$$

Which upon applying the monotone convergence theorem twice (once for each integral), in conjunction with the previous lemmas, gives

$$\mu\left(\bigcup_n A_n\right) = \lim_{N \rightarrow \infty} \sum_{n \leq N} \int_{E_1} \left(\int_{E_2} 1_{A_n}(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1) = \sum_{n=1}^{\infty} \mu(A_n)$$

\square

Theorem (Fubini's Theorem). *Let $(E, \mathcal{E}, \mu) = (E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2, \mu_1 \otimes \mu_2)$. Then*

(a) *Let $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ be measurable and non-negative. Then*

$$\mu(f) = \int_E f d\mu = \int_{E_1} \left(\int_{E_2} f(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1) \quad (\dagger)$$

$$= \int_{E_2} \left(\int_{E_1} f(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2) \quad (\circ)$$

(b) *If $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ is μ -integrable, then if*

$$A_1 = \left\{ x_1 \in E_1 : \int_{E_2} |f(x_1, x_2)| d\mu_2(x_2) < \infty \right\}$$

and for $f_1(x_1) = \int_{E_2} f(x_1, x_2) d\mu_2(x_2)$ for $x_1 \in A_1$, and $f_1(x_1) = 0$ on A_1^c , we have $\mu_1(A_1^c) = 0$, and $\mu(f) = \mu_1(f_1) = \mu_1(f_1 1_{A_1})$.

Remark: in (b), if f is bounded, $A_1 = E_1$. The same statement holds for f_2 , A_2 with the obvious modifications in (b), so $\mu_1(f_1) = \mu_2(f_2)$. But for $f(x_1, x_2) = \frac{x_1^2 - x_2^2}{(x_1^2 + x_2^2)}$ on $(0, 1)^2$, we have $\mu_1(f_1) \neq \mu_2(f_2)$ but f is not Lebesgue measurable on $(0, 1)^2$.

Proof. By the construction of $\mu(A)$ for rectangles $A = A_1 \times A_2 \in \mathcal{A}$ generating \mathcal{E} , the identities (\dagger) and (\circ) hold for $f = 1_A$, and by uniqueness of extension, this extends to 1_A , $A \in \mathcal{E}$, and by linearity of the integral this extends to simple functions. By the monotone convergence theorem (applied 5 times) on simple functions $0 \leq f_n \uparrow f$, the result (a) follows.

If $h(x_1) = \int_{E_2} |f(x_1, x_2)| d\mu_2(x_2)$, then by (a) $\mu_1(|h|) \leq \mu(|f|) < \infty$ since f is μ -integrable. So f_1 is μ_1 -integrable and $\mu_1(A_1^c) = 0$. Then $f_1^\pm = \int_{E_2} f^\pm(x_1, x_2) d\mu_2(x_2)$ so $\mu_1(f_1) = \mu_1(f_1^+) - \mu_1(f_1^-)$. Thus

$$\mu(f) = \mu(f^+) - \mu(f^-) = \mu_1(f_1^+) - \mu_1(f_1^-) = \mu_1(f_1)$$

by (a). □

Remark: the preceding results for product measures extend to σ -finite measures μ .

For (E_i, \mathcal{E}_i) for $i = 1, \dots, n$ with σ -finite μ_i , then since

$$(\mathcal{E}_1 \otimes \mathcal{E}_2) \otimes \mathcal{E}_3 = \mathcal{E}_1 \otimes (\mathcal{E}_2 \otimes \mathcal{E}_3)$$

by a π -system argument and Dynkin's lemma, we can iterate the construction of product measures to obtain $\mu_1 \otimes \dots \otimes \mu_n$, a unique product measure on $(\bigotimes_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{E}_i)$ such that $\mu_1 \otimes \dots \otimes \mu_n(A_1 \times \dots \times A_n) = \prod_{i=1}^n \mu_i(A_i)$.

In particular, on \mathbb{R}^n with Borel- σ -algebra $\mathcal{B}(\mathbb{R}^n) = \bigotimes_{i=1}^n \mathcal{B}(\mathbb{R})$ (product topology), we obtain the n -dimensional Lebesgue measure

$$\mu^n = \bigotimes_{i=1}^n \mu$$

and Fubini's theorem (applied $n - 1$ times) implies

$$\mu^n(f) = \int_{\mathbb{R}^n} f d\mu^n = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(x_1, \dots, x_n) d\mu(x_1) \dots d\mu(x_n)$$

whenever f is measurable and non-negative, or μ^n -integrable.

Product Probability Spaces & Independence

Proposition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(E, \mathcal{E}) = (\bigoplus_{i=1}^n E_i, \bigoplus_{i=1}^n \mathcal{E}_i)$. Consider $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ measurable and such that $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$. The following are equivalent:

- (i) X_1, \dots, X_n are independent
- (ii) $\mu_X = \bigoplus_{i=1}^n \mu_{X_i}$
- (iii) For all $f_i : E_i \rightarrow \mathbb{R}$ bounded and measurable,

$$\mathbb{E} \left[\prod_{i=1}^n f_i(X_i) \right] = \prod_{i=1}^n \mathbb{E} f_i(X_i)$$

Proof. First we show (i) implies (ii): for rectangles $A = \times_{i=1}^n A_i$, $A_i \in \mathcal{E}_i$, we have (by the definition of image measure)

$$\mu_X(A_1, \dots, A_n) = \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i) = \prod_{i=1}^n \mu_{X_i}(A_i)$$

Now we show (ii) implies (iii): by Fubini's theorem,

$$\begin{aligned} \mathbb{E} \left[\prod_{i=1}^n f_i(X_i) \right] &= \mu_X \left(\prod_{i=1}^n f_i(X_i) \right) = \int_{E_1} \dots \int_{E_n} f_1(x_1) \dots f_n(x_n) d\mu_{X_1}(x_1) \dots d\mu_{X_n}(x_n) \\ &= \prod_{i=1}^n \int_{E_i} f_i(x_i) d\mu_{X_i}(x_i) = \prod_{i=1}^n \mathbb{E} f_i(X_i) \end{aligned}$$

Finally we show (iii) implies (i): take $f_i = 1_{A_i}$ for any $A_i \in \mathcal{E}_i$, which is bounded and measurable. So

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{E} \left[\prod_{i=1}^n 1_{A_i}(X_i) \right] = \prod_{i=1}^n \mathbb{E} 1_{A_i} = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)$$

so X_1, \dots, X_n are independent. \square

4 L^p -spaces and norms

Definition. A *norm* on a vector space V (over \mathbb{R}) is a map $\|\cdot\|_V : V \rightarrow \mathbb{R}_+$ such that

1. $\|\lambda v\| = |\lambda| \cdot \|v\|$
2. $\|u + v\| \leq \|u\| + \|v\|$
3. $\|v\| = 0 \iff v = 0$

Definition. For (E, \mathcal{E}, μ) a measure space, we define $L^p(E, \mathcal{E}, \mu) = L^p(\mu) = L^p$ by

$$L^p(E, \mathcal{E}, \mu) = \{f : E \rightarrow \mathbb{R} \text{ measurable s.t. } \|f\|_p < \infty\}$$

where

$$\|f\|_p = \left(\int_E |f(x)|^p d\mu(x) \right)^{1/p}, \quad 1 \leq p < \infty$$

$$\|f\|_\infty = \text{ess sup } |f| := \inf\{\lambda > 0 : |f| \leq \lambda \text{ a.e.}\}$$

The property (1) of a norm holds for $\|\cdot\|_p$ whenever $1 \leq p \leq \infty$. Property (2) holds for $p = 1, \infty$ and also for $1 < p < \infty$ (to be proved). For (3), note that $f = 0$ implies $\|f\|_p = 0$, but $\|f\|_p = 0$ implies $f = 0$ almost everywhere on E . We can define quotient spaces

$$\mathcal{L}_p = L^p / \{f = 0 \text{ a.e.}\} = \{[f] : f \in L^p\}$$

where the equivalence classes are $[f] = \{g \in L^p : g = f \text{ a.e.}\}$. The functional $\|\cdot\|_p$ is then a norm on \mathcal{L}_p .

Proposition (Chebyshev's/Markov's inequality). Let $f \geq 0$ be non-negative and measurable. Then for all $\lambda > 0$, $\mu(f \geq \lambda) = \mu(\{x : f(x) \geq \lambda\}) \leq \frac{\mu(f)}{\lambda}$.

Proof. Integrate $\lambda 1_{\{f \geq \lambda\}} \leq f$ on E . □

Definition. Let $I \subseteq \mathbb{R}$ be an interval, then a map $c : I \rightarrow \mathbb{R}$ is called *convex* if

$$c(tx + (1-t)y) \leq tc(x) + (1-t)c(y), \quad \forall x, y \in I, \quad \forall t \in (0, 1)$$

which is easily seen to be equivalent to the condition that for all $x, y \in I$ and t with $x < t < y$,

$$\frac{c(t) - c(x)}{t - x} \leq \frac{c(y) - c(t)}{y - t} \quad (\circ)$$

Since c is continuous on the interior of I , it is Borel-measurable.

Lemma. Let $m \in \text{int}(I)$. Then if c is convex on I , there exist a, b such that $c(x) \geq ax + b$ with equality when $x = m$.

Proof. Define

$$a = \sup\left\{\frac{c(m) - c(x)}{m - x} : x < m\right\}$$

which exists in \mathbb{R} by (o). Let $y \in I$, $y > m$, then by (o), $a \leq \frac{c(y) - c(m)}{y - m}$, so we get

$$c(y) \geq ay + \underbrace{(-am + c(m))}_{=b}$$

Likewise for $x < m$, by definition of a

$$\frac{c(m) - c(y)}{m - y} \leq a$$

so $c(y) \geq ay - b$. Also $c(m) = am + b$. □

Theorem (Jensen's inequality). *Let X be a random variable taking values in $I \subseteq \mathbb{R}$ and such that $\mathbb{E}|X| < \infty$. If $c : I \rightarrow \mathbb{R}$ is convex, then $\mathbb{E}c(X) \geq c(\mathbb{E}X)$, in particular $\mathbb{E}c(X) = \mathbb{E}c^+(X) - \mathbb{E}c^-(X)$ is well defined in $(-\infty, \infty]$.*

Proof. Define $m = \mathbb{E}X = \int_I z d\mu_X(z)$, and if $m \notin \text{int}(I)$, then $X = m$ almost surely and the result follows. If $m \in \text{int}(I)$, then we can apply the lemma to see $c^-(X) \leq |a||X| + |b|$. So $\mathbb{E}c^-(X) \leq |a|\mathbb{E}|X| + |b| < \infty$, and $\mathbb{E}c(X) = \mathbb{E}c^+(X) - \mathbb{E}c^-(X)$ is well-defined in $(-\infty, \infty]$.

Then integrating the inequality from the lemma

$$\mathbb{E}c(X) \geq a\mathbb{E}X + b = am + b = c(m) = c(\mathbb{E}X)$$

□

Remark: a consequence of this is that if X is a bounded random variable (in $L^\infty(\mathbb{P})$), and if $1 \leq p < q < \infty$ then $c(x) = |x|^{q/p}$ is convex and

$$\|X\|_p = (\mathbb{E}|X|^p)^{1/p} = c(\mathbb{E}|X|^p)^{1/q} \leq \mathbb{E}(c(|X|^p))^{1/q} = \|X\|_q$$

Using the monotone convergence theorem, this extends to all $X \in L^q(\mathbb{P})$. In particular $L^q(\mathbb{P}) \subseteq L^p(\mathbb{P})$ for all $1 \leq p \leq q \leq \infty$.

Theorem (Holders inequality). *Let f, g be measurable on (E, \mathcal{E}, μ) . If p, q are conjugate, i.e. $\frac{1}{p} + \frac{1}{q} = 1$ and $1 \leq p, q \leq \infty$, then*

$$\mu(|fg|) = \int_E |fg| d\mu \leq \|f\|_p \|g\|_q$$

(for $p = q = 2$, this is the Cauchy-Schwarz inequality on L^2)

Proof. The cases $p = 1, \infty$ are obvious, and we can assume $f \in L^p, g \in L^q$ (or else we're done). We can also assume that we don't have $f = 0$ almost everywhere (else done), hence $\|f\|_p > 0$, so by dividing we can assume $\|f\|_p = 1$. Then

$$\mu(|fg|) = \int_E |g| \frac{1}{|f|^{p-1}} 1_{\{|f|>0\}} \underbrace{|f|^p}_{d\mathbb{P}} d\mu \leq \left(\int_E |g|^q \frac{1}{|f|^{q(p-1)}} |f|^p d\mu \right)^{1/q} = \|g\|_q$$

□

Theorem (Minkowski's inequality). *Let $f, g : (E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ be measurable. Then for all $1 \leq p \leq \infty$*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

Proof. $p = 1, \infty$ are clear, so assume $1 < p < \infty$. We may assume $f, g \in L^p$ or else it is obvious. We can integrate the pointwise inequality

$$|f + g|^p \leq 2^p(|f|^p + |g|^p)$$

to deduce

$$\|f + g\|_p^p \leq 2^p (\|f\|_p^p + \|g\|_p^p) < \infty$$

So we can assume $0 < \|f + g\|_0 < \infty$. Now

$$\|f + g\|_p^p = \int_E |f + g|^{p-1} |f + g| d\mu = \int_E |f + g|^{p-1} |f| d\mu + \int_E |f + g|^{p-1} |g| d\mu$$

So by Holders inequality with q conjugate to p

$$\|f + g\|_p^p \leq \underbrace{\left(\int_E |f + g|^{q(p-1)} d\mu \right)^{1/q}}_{\|f + g\|_p^{p/q}} (\|f\|_p + \|g\|_p)$$

So obtain $\|f + g\|_p \leq \|f\|_p + \|g\|_p$. \square

Theorem (\mathcal{L}^p is a Banach space). *Let $1 \leq p \leq \infty$, and let $f_n \in L^p$ be a Cauchy sequence. Then there exists $f \in L^p$ such that $f_n \rightarrow f$ in L^p .*

Proof. We assume $p < \infty$, the proof when $p = \infty$ is easier. or all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\forall m, n \geq N$, $\|f_n - f_m\| \leq \varepsilon$. Using this with $\varepsilon = 2^{-k}$ we can extract a subsequence f_{N_k} such that $S = \sum_{k=1}^{\infty} \|f_{N_{k+1}} - f_{N_k}\|_p \leq \sum_{k=1}^{\infty} 2^{-k} < \infty$. By Minkowski's inequality, for any K

$$\left\| \sum_{k=1}^K |f_{N_{k+1}} - f_{N_k}| \right\|_p \leq \sum_{k=1}^K \|f_{N_{k+1}} - f_{N_k}\|_p \leq S$$

So by the monotone convergence theorem applied to $\left| \sum_{k=1}^K |f_{N_{k+1}} - f_{N_k}| \right|^p \uparrow \left| \sum_{k=1}^{\infty} |f_{N_{k+1}} - f_{N_k}| \right|^p$ we see that

$$\left\| \sum_{k=1}^{\infty} |f_{N_{k+1}} - f_{N_k}| \right\|_p \leq S < \infty$$

Since the integral is finite, we see that $\sum_{k=1}^{\infty} |f_{N_{k+1}} - f_{N_k}| < \infty$ almost everywhere. Then $\sum_{k=1}^K (f_{N_{k+1}}(x) - f_{N_k}(x)) = f_{N_{K+1}}(x) - f_{N_1}(x)$ converges in \mathbb{R} for

all x in some set A with $\mu(A^c) = 0$. Since \mathbb{R} is complete, $f_{N_k}(x)$ converges in \mathbb{R} and we define

$$f(x) = \begin{cases} \lim_{k \rightarrow \infty} f_{N_k}(x) & x \in A \\ 0 & x \notin A \end{cases}$$

So $f_{N_k} \rightarrow f$ as $k \rightarrow \infty$ almost everywhere. Next

$$\|f_n - f\|_p^p = \mu(|f_n - f|^p) = \mu\left(\lim_k |f_n - f_{N_k}|^p\right) \leq \liminf_k \mu(|f_n - f_{N_k}|^p)$$

where the last inequality follows from Fatou's lemma. Using the Cauchy property, $\|f\|_p \leq \|f - f_N\|_p + \|f_N\|_p < \infty$, so $f \in L^p$ and $\|f_n - f_{N_k}\|_p^p \leq \varepsilon^p$ for $n, N_k \geq N$, so $f_n \rightarrow f$ in L^p . \square

Remark: if V is any of the spaces $C([a, b])$, $\{f \text{ simple}\}$ or $\{f \text{ a linear combination of indicators of intervals}\}$, then V is dense in $L^1(\mu)$, for μ the Lebesgue measure on $\mathcal{B}([a, b])$, and so the completion $(V, \|\cdot\|_1) = L^1(\mu)$.

$\mathcal{L}^2(\mu)$ as a Hilbert space

Definition. A symmetric bilinear form $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ on a vector space V is called a *inner product* if $\langle v, v \rangle \geq 0$ with equality only when $v = 0$. In this case we can define a norm $\|v\| = \sqrt{\langle v, v \rangle}$ on V , and if $(V, \langle \cdot, \cdot \rangle)$ is complete for $\|\cdot\|$, we call this space a *Hilbert space*.

Corollary. $\mathcal{L}^2(\mu)$ is a Hilbert space for $\langle f, g \rangle = \int_E f g d\mu$.

Proof. Trivial by previous theorem. \square

Pythagoras rule: for $f, g \in L^2$, $\|f + g\|_2^2 = \|f\|_2^2 + 2\langle f, g \rangle + \|g\|_2^2$.

We say that f is orthogonal to g if $\langle f, g \rangle = \int_E f g d\mu = 0$, and write $f \perp g$. For centred (mean 0) random variables X, Y , we have $\langle X, Y \rangle = \mathbb{E}(XY) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \text{Cov}(X, Y) = 0$ whenever $X \perp Y$.

Parallelogram identity: $\|f + g\|_2^2 + \|f - g\|_2^2 = 2(\|f\|_2^2 + \|g\|_2^2)$

For $V \subseteq \mathcal{L}^2(\mu)$, we define its orthogonal complement

$$V^\perp = \{f \in L^2(\mu) : \langle f, v \rangle = 0 \forall v \in V\}$$

We say that a subset V of \mathcal{L}^2 is closed if for any sequence f_n in V , which converges to some $f \in \mathcal{L}^2$, we have $f = v$ almost everywhere for some $v \in V$.

Theorem. Let V be a closed linear subspace of $\mathcal{L}^2(\mu)$. Then for all $f \in \mathcal{L}^2$ there exists a decomposition $f = v + u$, for $v \in V$, $u \in V^\perp$ such that $\|f - v\|_2 \leq \|f - g\|_2$ for all $g \in V$, with equality only if $g = v$ almost everywhere. We call v the projection of f onto V .

Proof. Define (throughout this proof we write $\|\cdot\|$ for $\|\cdot\|_2$) $d(f, V) = \inf_{g \in V} \|g - f\|$ and take $g_n \in V$ approximating the infimum. By the parallelogram-law

$$\begin{aligned} 2\|f - g_n\|^2 + 2\|f - g_m\|^2 &= \|2f - (g_n + g_m)\|^2 + \|g_n - g_m\|^2 \\ &= 4 \left\| f - \underbrace{\frac{g_n + g_m}{2}}_{\in V} \right\|^2 + \|g_n - g_m\|^2 \\ &\geq 4d(f, V)^2 + \|g_n - g_m\|^2 \end{aligned}$$

So $\limsup_{m,n} \|g_n - g_m\|^2 \leq 4d(f, V)^2 - 4d(f, V)^2 = 0$. So (g_n) is Cauchy in L^2 , so by completeness, it converges $g_n \rightarrow v$ for some $v \in L^2$, and since V is closed, $v \in V$. In particular, $\inf_{g \in V} \|g - f\| = \|v - f\|$.

We further have

$$d(f, V)^2 \leq F(t) := \|f - (v + th)\|^2, \quad t \in \mathbb{R}, \quad h \in V$$

from which we obtain the first order condition $F'(0) = 2\langle f - v, h \rangle = 0$ for all $h \in V$. So if we define $f - v = u$, we have $f = u + v$ and $u \in V^\perp$ since h was arbitrary. If $f = w + z$ with $w \in V$, $z \in V^\perp$ then

$$v - w + u - z = f - f = 0$$

so $\|v - w + u - z\|^2 = 0$ with $v - w \in V$, $u - z \in V^\perp$ so by Pythagoras $\|v - w + u - z\|^2 = 0 = \|v - w\|^2 + \|u - z\|^2$, i.e $v = w$ and $u = z$ (almost everywhere). \square

Convergence in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ and uniform integrability (UI)

Theorem (Bounded convergence). *Let X_n be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|X_n| \leq C < \infty$ for all n , and $X_n \xrightarrow{\mathbb{P}} X$ as $n \rightarrow \infty$. Then $X_n \rightarrow X$ in $L^1(\mathbb{P})$.*

Proof. We know $X_{n_k} \rightarrow X$ almost surely along a subsequence, so

$$|X| = \text{a.s.} \lim_k |X_{n_k}| \leq C$$

so X is also bounded by C . Then

$$\mathbb{E}|X_n - X| (1_{|X_n - X| > \varepsilon/2} + 1_{|X_n - X| \leq \varepsilon/2}) \leq 2C\mathbb{P}(|X_n - X| > \varepsilon/2) + \varepsilon/2$$

Which is less than ε for all n sufficiently large. \square

If $X \in L^1(\mathbb{P})$, then on $\delta \rightarrow 0$,

$$I_X(\delta) := \sup \{ \mathbb{E}(|X|1_A) : A \in \mathcal{F}, \mathbb{P}(A) \leq \delta \} \downarrow 0$$

Suppose not, then there exists $\varepsilon > 0$ and $A_n \in \mathcal{F}$ such that $\mathbb{P}(A_n) \leq 2^{-n}$ but $\mathbb{E}(|X|1_{A_n}) \geq \varepsilon > 0$ for all n .

Since $\sum_n \mathbb{P}(A_n) < \infty$, we use the Borel-Cantelli lemma to see

$$\mathbb{P}\left(\bigcap_n \bigcup_{m \geq n} A_m\right) = 0$$

But $\mathbb{E}(|X|1_{A_n}) \leq \mathbb{E}(|X|1_{\bigcup_{m \geq n} A_m})$ and noting that

$1_{\left(\bigcup_{m \geq n} A_m\right)} \rightarrow 1_{\left(\bigcap_n \bigcup_{m \geq n} A_m\right)}$, we have that $\mathbb{E}|X|1_{\bigcup_{m \geq n} A_m} \rightarrow \mathbb{E}|X|1_{\bigcap_n \bigcup_{m \geq n} A_m}$ by the dominated convergence theorem with dominating function $g(x) = |X|1_\Omega$. So $\mathbb{E}|X|1_{A_n} \rightarrow 0$, a contradiction.

For a collection $\mathcal{X} \subseteq L^1(\mathbb{P})$ of random variables, we say \mathcal{X} is *uniformly integrable* if it is bounded in $L^1(\mathbb{P})$ and

$$I_{\mathcal{X}}(\delta) = \sup \{ \mathbb{E}|X|1_A : A \in \mathcal{F}, \mathbb{P}(A) \leq \delta, X \in \mathcal{X} \} \downarrow 0$$

as $\delta \rightarrow 0$. Note that $X_n = n1_{(0,1/n]}$ for μ Lebesgue measure on $(0,1)$ is bounded in $L^1(\mathbb{P})$ but not uniformly integrable. If \mathcal{X} is bounded in $L^p(\mathbb{P})$ for $p > 1$, then by Holder's inequality

$$\mathbb{E}|X|1_A \leq \underbrace{\|X\|_p}_{\leq C} \underbrace{\mathbb{P}(A)^{1/q}}_{\leq \delta^{1/q}} \rightarrow 0 \text{ uniformly}$$

so such \mathcal{X} is uniformly integrable.

Lemma. $\mathcal{X} \subseteq L^1(\mathbb{P})$ is uniformly integrable if and only if $\sup_{X \in \mathcal{X}} \mathbb{E}(|X|1_{\{|X|>k\}}) \rightarrow 0$ as $k \rightarrow \infty$.

Proof. If \mathcal{X} is uniformly integrable, then by Markov's inequality

$$\mathbb{P}(|X| > k) \leq \frac{\mathbb{E}|X|1_\Omega}{k} \leq \frac{I_{\mathcal{X}}(1)}{k} \rightarrow 0 \text{ as } k \rightarrow \infty$$

so $\mathbb{P}(|X| > k) < \delta$ uniformly for k sufficiently large. So using the uniformly integrable property with $A = \{|X| > k\}$ we get the required limit.

Conversely, we have $\mathbb{E}|X| = \mathbb{E}|X|(1_{\{|X| \leq k\}} + 1_{\{|X| > k\}}) \leq k + \varepsilon/2$ for k large enough, so \mathcal{X} is bounded in $L^1(\mathbb{P})$. Next for A such that $\mathbb{P}(A) \leq \delta$

$$\mathbb{E}|X|1_A (1_{\{|X| \leq k\}} + 1_{\{|X| > k\}}) \leq k\mathbb{P}(A) + \mathbb{E}|X|1_{\{|X| > k\}} \leq k\delta + \varepsilon/2$$

for k large. □

Theorem. Let X_n, X be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the following are equivalent

1. $X_n, X \in L^1(\mathbb{P})$, $X_n \rightarrow X$ in $L^1(\mathbb{P})$ as $n \rightarrow \infty$.

2. $(X_n : n \in \mathbb{N})$ is uniformly integrable and $X_n \xrightarrow{\mathbb{P}} X$ as $n \rightarrow \infty$.

Proof. We first show (1) implies (2): clearly $\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}|X_n - X|}{\varepsilon} \rightarrow 0$, so $X_n \rightarrow X$ in probability. Since any finite collection is uniformly integrable, so are X_1, \dots, X_N , and for $n \geq N$ and A with $\mathbb{P}(A) \leq \delta$ we have $\mathbb{E}|X_n|1_A \leq \mathbb{E}|X_n - X|1_A + \mathbb{E}|X|1_A \leq \varepsilon/2 + \varepsilon/2$ for N large enough and δ small enough.

Now we show (2) implies (1). Since $X_{n_k} \rightarrow X$ almost surely along a subsequence,

$$\mathbb{E}|X| = \mathbb{E} \liminf_k |X_{n_k}| \leq \liminf_k \mathbb{E}|X_{n_k}| \leq I_{\mathcal{X}}(1) < \infty$$

So $X \in L^1(\mathbb{P})$. Next define

$$X_n^K = \min(\max(-k, X_n), k) = g(X_n)$$

$$X^K = \min(\max(-k, X), k) = g(X)$$

Then

$$\mathbb{P}(|g(X_n) - g(X)| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon') \text{ for some } \varepsilon' > 0$$

And since $X_n \xrightarrow{\mathbb{P}} X$, the RHS converges to 0 as $n \rightarrow \infty$. Now by bounded convergence, $X_n^K \rightarrow X^K$ in $L^1(\mathbb{P})$ and so

$$\mathbb{E}|X_n - X| \leq \underbrace{\mathbb{E}|X_n - X_n^K|}_{\mathbb{E}|X_n|1_{\{|X_n|>K\}} \rightarrow 0} + \mathbb{E}|X_n^K - X^K| + \underbrace{\mathbb{E}|X^K - X|}_{\mathbb{E}|X|1_{\{|X|>K\}} \rightarrow 0} < \varepsilon$$

□

Fourier transforms

In this section, we will write $L^p(\mathbb{R}^d)$ for the set of measurable functions $f : \mathbb{R}^d \rightarrow \mathbb{C}$ such that $\|f\|_p = \left(\int_{\mathbb{R}^d} |f(x)|^p dx\right)^{1/p} < \infty$.

We can extend the integral as a complex linear map $L^1(\mathbb{R}^d) \rightarrow \mathbb{C}$ by $\int_{\mathbb{R}^d} (u + iv)(x) dx := \int_{\mathbb{R}^d} u(x) dx + i \int_{\mathbb{R}^d} v(x) dx$.

Note that

$$\left\| \int_{\mathbb{R}^d} f(x) dx \right\| = \int_{\mathbb{R}^d} \alpha f(x) dx$$

for some $\alpha \in \mathbb{C}$ with $|\alpha| = 1$ (take α to have complementary argument to the integral). Then write $\alpha f = u + iv$ so

$$\begin{aligned} \left\| \int_{\mathbb{R}^d} f(x) dx \right\| &= \int_{\mathbb{R}^d} u(x) dx + i \underbrace{\int_{\mathbb{R}^d} v(x) dx}_{=0} \\ &\leq \int_{\mathbb{R}^d} |f(x)| dx \end{aligned}$$

since $u \leq |u| \leq |\alpha f| = |f|$.

For $f \in L^1(\mathbb{R}^d)$ we define the *Fourier transform* \hat{f} by

$$\hat{f}(u) = \int_{\mathbb{R}^d} f(x) e^{i\langle u, x \rangle} dx$$

Where $\langle u, x \rangle$ is the usual dot product on \mathbb{R}^d , i.e $\sum_{i=1}^d u_i x_i$.

Note that $|\hat{f}(u)| \leq \|f\|_1$. Also, if $u_n \rightarrow u$, then $e^{i\langle u_n, x \rangle} \rightarrow e^{i\langle u, x \rangle}$, so by the dominated convergence theorem (dominated by $|f|$), $\hat{f}(u_n) \rightarrow \hat{f}(u)$. So \hat{f} is a continuous bounded function.

For $f \in L^1(\mathbb{R}^d)$, with $\hat{f} \in L^1(\mathbb{R}^d)$, we say that the *Fourier inversion formula* holds for f if

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(u) e^{-i\langle u, x \rangle} du$$

for almost all $x \in \mathbb{R}^d$.

For $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ (neither L^1 nor L^2 is contained in the other with the Lebesgue measure), we say the *Plancherel identity* holds if $\|\hat{f}\|_2 = (2\pi)^{d/2} \|f\|_2$.

We'll show the inversion formula holds whenever $\hat{f} \in L^1$, and Plancherel holds for all $f \in L^1 \cap L^2$.

For a finite Borel-measure μ on \mathbb{R}^d , we define the *Fourier transform* $\hat{\mu}$ by

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} d\mu(x), \quad u \in \mathbb{R}^d$$

Then $|\hat{\mu}| \leq \mu(\mathbb{R}^d)$. If μ has a density f with respect to the Lebesgue measure dx , then $\hat{\mu} = \hat{f}$.

For a random variable X in \mathbb{R}^d , the *characteristic function* ϕ_X is given by $\phi_X(u) = \mathbb{E}(e^{i\langle u, X \rangle}) = \mu_X(u)$, where μ_X is the Law of X .

Convolutions

For $f \in L^1(\mathbb{R}^d)$ and a probability measure ν on \mathbb{R}^d , we define the convolution $f * \nu$ by

$$f * \nu(x) = \begin{cases} \int_{\mathbb{R}^d} f(x-y) d\nu(y) & \text{if } f(x-\cdot) \in L^1(\nu) \\ 0 & \text{otherwise} \end{cases}$$

Note that for $p \in [1, \infty)$

$$\begin{aligned} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-y)| d\nu(y) \right)^p dx &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x-y)|^p d\nu(y) dx \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x)|^p dx d\nu(x) \\ &= \|f\|_p^p \end{aligned}$$

Where the inequality follows from Jensen, the swapping of integration from Fubini and using translation invariance of the Lebesgue measure.

So for $f \in L^p$, have $f(x-\cdot) \in L^p(\nu)$ at almost all x , and

$$\|f * \nu\|_p^p = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x-y) d\nu(y) \right|^p dx \leq \|f\|_p^p$$

So $f \mapsto f * \nu$ is a contraction on $L^p(\mathbb{R}^d)$. In the case ν has a density g with respect to the Lebesgue measure dx , we write $f * g = f * \nu$.

For probability measures μ, ν on \mathbb{R}^d , we define $\mu * \nu$, another probability measure on \mathbb{R}^d as the law of $X + Y$ where X, Y are independent with laws μ, ν . i.e

$$\begin{aligned} \mu * \nu(A) &= \mathbb{P}(X + Y \in A) = \int_{\mathbb{R}^d \times \mathbb{R}^d} 1_A(x+y) d(\mu \otimes \nu)(x, y) \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} 1_A(x+y) d\mu(x) \right) d\nu(y) \end{aligned}$$

by Fubini. If μ has density f with respect to dx , then $\mu * \nu$ has density $f * \nu$ with respect to dx .

Indeed

$$\begin{aligned}
 \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} 1_A(x+y) d\mu(x) \right) d\nu(y) &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} 1_A(x+y) f(x) dx \right) d\nu(y) \\
 &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} 1_A(v) f(v-y) dv \right) d\nu(y) \\
 &= \int_{\mathbb{R}^d} 1_A(v) \underbrace{\int_{\mathbb{R}^d} f(v-y) d\nu(y)}_{=f*\nu(v) \text{ a.e}} dv
 \end{aligned}$$

Exercise: $\hat{f * \nu}(u) = \hat{f}(u) \hat{\nu}(u)$

Fourier transforms of Gaussians

The pdf of a $\mathcal{N}(0, t)$, $t > 0$ variable on \mathbb{R} is

$$g_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}, \quad x \in \mathbb{R}$$

If ϕ_X is the characteristic function of $X \sim \mathcal{N}(0, 1)$, then

$$\frac{d}{du} \phi_X(u) = \frac{d}{du} \int_{\mathbb{R}} e^{iux} g_1(x) dx$$

Which by the Theorem on differentiation under the integral sign is equal to

$$\int_{\mathbb{R}} \left[\frac{d}{du} e^{iux} \right] g_1(x) dx = i \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iux} x e^{-x^2/2} dx$$

And by integration by parts on $v = e^{iux}$ and $w' = x e^{-x^2/2}$ this is

$$i^2 \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u e^{iux} e^{-x^2/2} dx = -u \phi_X(u)$$

So this implies

$$\frac{d}{du} \left(e^{u^2/2} \phi_X(u) \right) = u e^{u^2/2} \phi_X(u) - e^{u^2/2} u \phi_X(u) = 0$$

So $\phi_X(u) = \phi_X(0) e^{-u^2/2} = e^{-u^2/2}$. In other words, we have

$$\hat{g}_1(u) = (2\pi)^{1/2} g_1(u)$$

In \mathbb{R}^d , consider a Gaussian random vector $Z = (Z_1, \dots, Z_d)$ with iid $\mathcal{N}(0, 1)$ components. Then the joint pdf of $\sqrt{t}Z$ is

$$g_t(x) = \prod_{j=1}^d \frac{1}{\sqrt{2\pi t}} e^{-\frac{x_j^2}{2t}} = (2\pi t)^{-d/2} e^{-\frac{|x|^2}{2t}}, \quad x \in \mathbb{R}^d$$

The Fourier transform of g_t equals

$$\begin{aligned} \hat{g}_t(u) &= \mathbb{E}[e^{i\langle u, \sqrt{t}Z \rangle}] = \mathbb{E} \prod_{j=1}^d e^{iu_j \sqrt{t} Z_j} \\ &= \prod_{j=1}^d \mathbb{E} e^{iu_j \sqrt{t} Z_j} = \prod_{j=1}^d \exp \left(-\frac{u_j^2 t}{2} \right) = \exp \left(-\frac{|u|^2 t}{2} \right) \end{aligned}$$

And so

$$\hat{g}_t(u) = (2\pi)^{d/2} t^{-d/2} g_{1/t}(u), \quad u \in \mathbb{R}^d$$

Taking Fourier transforms now with respect to u

$$\hat{\hat{g}}_t = (2\pi)^d g$$

And since g_t is an even function, and since dx is translation invariant (TIV) we see

$$g(x) = (2\pi)^{-d} \hat{g}_t(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \hat{g}_t(u) du$$

So Fourier inversion holds for (such) Gaussians.

We say that a function on \mathbb{R}^d is a Gaussian convolution if it is of the form

$$f * g_t(x) = \int_{\mathbb{R}^d} f(x-y) g_t(y) dy, \quad x \in \mathbb{R}^d, \quad t > 0, \quad f \in L^1(\mathbb{R}^d)$$

One shows that $f * g_t$ is continuous on \mathbb{R}^d , $\|f * g_t\|_1 \leq \|f\|_1$, and $\widehat{f * g_t}(u) = \hat{f}(u) \cdot e^{-\frac{|u|^2 t}{2}}$ so $\|\widehat{f * g_t}\|_\infty \leq \|f\|_1$, $\|\widehat{f * g_t}\|_1 \leq \|f\|_1 (2\pi)^{d/2} t^{-d/2} < \infty$.

Lemma. *Fourier inversion holds for all Gaussian convolutions.*

Proof. We use Fourier inversion for $g_t(y)$ to see

$$(2\pi)^d f * g_t(x) = (2\pi)^d \int_{\mathbb{R}^d} f(x-y) g_t(y) dy = \int_{\mathbb{R}^d} f(x-y) \int_{\mathbb{R}^d} e^{-i\langle u, y \rangle} \hat{g}_t(u) du dy$$

Applying Fubini (check it is integrable in product measure) we get

$$\begin{aligned} & \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \underbrace{\int_{\mathbb{R}^d} f(x-y) e^{i\langle u, x-y \rangle} dy}_{=\hat{f}(u)} \hat{g}_t(u) du \\ &= \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \widehat{f * g_t}(u) du \end{aligned}$$

□

Remark: if μ is a finite measure, then $\mu * g_t = \underbrace{\mu * g_{t/2} * g_{t/2}}_{\in L^1}$ so this is also a

Gaussian convolution.

Lemma (Gaussian convolutions are dense in L^p). *Let $f \in L^p$, $1 \leq p < \infty$, then $\|f * g_t - f\|_p \rightarrow 0$ as $t \rightarrow 0$.*

Proof. One shows that $C_c(\mathbb{R}^d)$ (the space of continuous functions with compact support) is dense in L^p (Example sheet). So for all $\varepsilon > 0$, there exists $h \in C_c(\mathbb{R}^d)$ such that $\|f - h\|_p < \varepsilon/3$, and by properties of $*$ we also have $\|f * g_t - h * g_t\|_p = \|(f - h) * g_t\|_p$, which by contraction properties of the convolution is $\leq \|f - h\|_p < \varepsilon/3$. So $\|f * g_t - f\|_p \leq \|f * g_t - h * g_t\|_p + \|h * g_t - h\|_p + \|h - f\|_p < \varepsilon$ for all t . So we can restrict to $f = h \in C_c(\mathbb{R}^d)$.

We define a new map $y \mapsto e(y) = \int_{\mathbb{R}^d} |h(x-y) - h(x)|^p dx$. Since h is bounded on its bounded support, the dominated convergence theorem implies that e is continuous at $y = 0$. So

$$\|h * g_t - h\|_p^p = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (h(x-y) - h(x)) g_t(y) dy \right|^p dx$$

Which by Jensen's inequality and Fubini is

$$\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |h(x-y) - h(x)|^p dx g_t(y) dy = \int_{\mathbb{R}^d} e(\sqrt{t}z) g_1(z) dz$$

Where $z = \frac{y}{\sqrt{t}}$. Now note $|e(y)| \leq 2^{p+1} \|h\|_p^p < \infty$ and $e(\sqrt{t}z) \rightarrow 0$ as $t \rightarrow 0$ pointwise so the dominated convergence theorem implies the RHS converges to 0 as $t \rightarrow 0$. \square

Theorem (Fourier inversion). *Let $f \in L^1(\mathbb{R}^d)$ be such that $\hat{f} \in L^1(\mathbb{R}^d)$. Then*

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle x, u \rangle} \hat{f}(u) du \quad \text{for almost all } x \in \mathbb{R}^d$$

Remarks:

- If $\hat{f} = \hat{g}$, then $\widehat{f - g} = 0 \in L^1$ so $f = g$ almost everywhere, and the Fourier transform is injective.
- The identity holds everywhere on \mathbb{R}^d for the (unique) continuous $f \in [f]$.

Proof. Since Fourier inversion holds for the Gaussian convolution

$$f * g_t(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \hat{f}(u) e^{-\frac{|u|^2 t}{2}} du = f_t(x)$$

Now by the previous lemma we have $f * g_t \xrightarrow{t \rightarrow 0} f$ in L^1 , so $f * g_t \rightarrow f$ in measure (Markov's inequality), and along a subsequence, $f * g_{t_n} \rightarrow f$ almost everywhere. On the other hand, by the dominated convergence theorem with dominating function $|\hat{f}|$, the RHS converges to $\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \hat{f}(u) du$, so this is equal to $\lim_{t_n \rightarrow 0} f_{t_n}$ almost everywhere by uniqueness of limits. \square

Theorem (Plancherel). *Let $f \in L^1 \cap L^2$. Then $\|f\|_2 = \frac{1}{(2\pi)^{d/2}} \|\hat{f}\|_2$.*

Remark: by Pythagoras' identity, $\langle f, g \rangle = \frac{1}{(2\pi)^d} \langle \hat{f}, \hat{g} \rangle$ for all $f, g \in L^1 \cap L^2$.

Proof. Initially, assume $\hat{f} \in L^1$ so $f, \hat{f} \in L^\infty$ (are bounded), and $(x, u) \mapsto f(x)\hat{f}(u)$ is $dx \otimes du$ integrable on $\mathbb{R}^d \times \mathbb{R}^d$, so Fubini's theorem for bounded functions applies. Hence

$$\begin{aligned} (2\pi)^d \|f\|_2^2 &= (2\pi)^d \int_{\mathbb{R}^d} f(x) \overline{f(x)} dx \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \hat{f}(u) du \right) \overline{f(x)} dx && \text{(Fourier inversion)} \\ &= \int_{\mathbb{R}^d} \hat{f}(u) \overline{\int_{\mathbb{R}^d} e^{i\langle u, x \rangle} f(x) dx} du && \text{(Fubini)} \\ &= \int_{\mathbb{R}^d} \hat{f}(u) \overline{\hat{f}(u)} du \\ &= \|\hat{f}\|_2^2 \end{aligned}$$

To extend this to general f , take Gaussian convolutions $f * g_t = f_t$ such that $f_t \rightarrow f$ in L^2 and by continuity of $\|\cdot\|_2$, this implies $\|f_t\|_2^2 \rightarrow \|f\|_2^2$. Also $|\hat{f}(u) e^{-\frac{|u|^2 t}{2}}|^2 \uparrow |\hat{f}(u)|^2$ so by monotone convergence $\|\hat{f}_t\|_2^2 \uparrow \|\hat{f}\|_2^2$. Therefore since the identity holds for $\widehat{f * g_t} \in L^1$, we see

$$\|f\|_2^2 = \lim_{t \rightarrow 0} \|f_t\|_2^2 = \lim_{t \rightarrow 0} \|\hat{f}_t\|_2^2 = \|\hat{f}\|_2^2$$

\square

Remark: since $L^1 \cap L^2$ is dense in L^2 we can extend the linear operator $F_0(f) = \frac{1}{(2\pi)^{d/2}} \hat{f}$ by continuity to a linear isometry $F : L^2 \rightarrow L^2$, known as the Fourier-Plancherel transform. One shows that F is surjective with inverse $F^{-1} : L^2 \rightarrow L^2$.

Fourier inversion (pointwise) for general probability measures on \mathbb{R}^d is more delicate. For the Dirac measure δ_0 on \mathbb{R} , with $\hat{\delta}_0 = \int_{\mathbb{R}} e^{iux} d\delta_0(x) = 1$ for all u . But the ‘inverse’ FT would be $\frac{1}{2\pi} \int_{\mathbb{R}} e^{iux} du$, which is not a Lebesgue integrable function.

Theorem. Let X be a random vector in \mathbb{R}^d with law μ_X . Then the characteristic function $\phi_X = \hat{\mu}_X$ uniquely determines μ_X . If $\phi_X \in L^1$ then μ_X has a probability density function $f_X(x)$ which is equal to $\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \phi_X(u) du$ almost everywhere.

Proof. Let $Z = (Z_1, \dots, Z_d)$ with iid entries $Z_j \sim \mathcal{N}(0, 1)$ independent of X . Then $\sqrt{t}Z$ has probability density function g_t and $X + \sqrt{t}Z$ has pdf $f_t = \mu_X * g_t$, a Gaussian convolution (equal to $\mu_X * g_{t/2} * g_{t/2}$), so

$$f_t(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} \phi_X(u) e^{-\frac{|u|^2 t}{2}} du$$

Which is uniquely determined by ϕ_X . We will use the fact (Example sheet) that two Borel probability measures μ, ν on \mathbb{R}^d coincide if and only if $\mu(g) = \nu(g)$ for all $g : \mathbb{R}^d \rightarrow \mathbb{R}$ bounded, continuous and compactly supported. Now

$$\int_{\mathbb{R}^d} g(x) f_t(x) dx = \mathbb{E}g(X + \sqrt{t}Z) \xrightarrow{t \rightarrow 0} \mathbb{E}g(X) = \int_{\mathbb{R}^d} g(x) d\mu_X(x) \quad (\dagger)$$

Since $|\cdot| \leq \|g\|_{\infty} < \infty$ and $X + \sqrt{t}Z \xrightarrow{\text{a.s.}} X$ as $t \rightarrow 0$ and bounded convergence. So by uniqueness of limits, ϕ_X determines μ_X (ϕ_X determines f_t).

Next, if $\phi_X \in L^1$, then by the dominated convergence theorem $f_t(x) \xrightarrow{t \rightarrow 0} f_X$ for some f_X for all $x \in \mathbb{R}^d$. In particular, since $\mu_X * g_t \geq 0$, the limit $f_X \geq 0$ on \mathbb{R}^d . Then for $g \in C_c^b(\mathbb{R}^d)$

$$\begin{aligned} \int_{\mathbb{R}^d} g(x) f_X(x) dx &= \int_{\mathbb{R}^d} g(x) \lim_{t \rightarrow 0} f_t(x) dx \\ &= \lim_{t \rightarrow 0} \int_{\mathbb{R}^d} g(x) f_t(x) dx && \text{(by dom conv)} \\ &= \int_{\mathbb{R}^d} g(x) d\mu_X(x) && \text{(by } \dagger) \end{aligned}$$

□

Definition. A sequence $(\mu_n : n \in \mathbb{N})$ of Borel probability measures on \mathbb{R}^d is said to *converge weakly* to a Borel probability measure μ on \mathbb{R}^d if

$$\mu_n(g) \xrightarrow{n \rightarrow \infty} \mu(g) \quad \forall g : \mathbb{R}^d \rightarrow \mathbb{R} \text{ bounded and continuous}$$

If X_n, X are random vectors with laws $\mu_{X_n} \rightarrow \mu_X$ weakly as $n \rightarrow \infty$, we say $X_n \rightarrow X$ weakly.

Remarks:

1. When $d = 1$, this is equivalent to $X_n \xrightarrow{d} X$ (Example sheet).
2. One shows that convergence of $\mu_n(g) \rightarrow \mu(g)$ for all $g \in C_c^\infty(\mathbb{R}^d)$ is sufficient to show $\mu_n \rightarrow \mu$ weakly.
3. This is equivalent to *weak-star* convergence in $(C_b(\mathbb{R}^d))^*$.

Theorem (Lery's continuity theorem). *If X_n, X are random vectors in \mathbb{R}^d , such that $\phi_{X_n}(u) \rightarrow \phi_X(u)$ for all $u \in \mathbb{R}^d$ as $n \rightarrow \infty$, then $\mu_{X_n} \rightarrow \mu_X$ weakly.*

Proof. Let $Z = (Z_1, \dots, Z_d)$ have iid $\mathcal{N}(0, 1)$ entries, independent of X_n, X . Any $g \in C_c^\infty(\mathbb{R}^d)$ is in $L^1(\mathbb{R}^d)$ and Lipschitz $|g(x) - g(y)| \leq \|g\|_{\text{Lip}}|x - y|$. For $\varepsilon > 0$, choose $t > 0$ small enough such that

$$\sqrt{t}\|g\|_{\text{Lip}}\mathbb{E}|Z| < \frac{\varepsilon}{3}$$

Then

$$\begin{aligned} |\mu_{X_n}(g) - \mu_X(g)| &= |\mathbb{E}g(X_n) - \mathbb{E}g(X)| \\ &\leq \mathbb{E}|g(X_n) - g(X_n + \sqrt{t}Z)| + \mathbb{E}|g(X) - g(X + \sqrt{t}Z)| \\ &\quad + |\mathbb{E}(g(X_n + \sqrt{t}Z) - g(X + \sqrt{t}Z))| \\ &\leq 2\|g\|_{\text{Lip}}\sqrt{t}\mathbb{E}|Z| + |\mathbb{E}(g(X_n + \sqrt{t}Z) - g(X + \sqrt{t}Z))| \\ &\leq 2\frac{\varepsilon}{3} + |\mathbb{E}(g(X_n + \sqrt{t}Z) - g(X + \sqrt{t}Z))| \end{aligned} \quad (*)$$

Now for the remaining term

$$\mathbb{E}g(X_n + \sqrt{t}Z) = \int_{\mathbb{R}^d} g(x)f_{t,n}(x)dx, \text{ where } f_{t,n} = g_t * \mu_{X_n}$$

So by Fourier inversion for Gaussian convolutions,

$$\mathbb{E}g(X_n + \sqrt{t}Z) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} g(x) \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \phi_{X_n}(u) e^{-\frac{|u|^2 t}{2}} du dx$$

Now with dominating function $|g(x)|e^{-\frac{|u|^2 t}{2}}$, which is $dx \otimes du$ integrable, the dominated convergence theorem implies for t fixed

$$\mathbb{E}g(X_n + \sqrt{t}Z) \rightarrow \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} g(x) \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} \phi_X(u) e^{-\frac{|u|^2 t}{2}} du dx = \int_{\mathbb{R}^d} g(x)f_t(x)dx$$

Where the last equality follows by Fourier inversion, with $f_t = g_t * \mu_X$. So the final integral equals $\mathbb{E}g(X + \sqrt{t}Z)$, and the final term in (*) is less than $\varepsilon/3$ for n large enough. \square

Theorem (Central limit theorem). *Let X_1, \dots, X_n be iid random variables with $\mathbb{E}X_i = 0$ for all i , and $\text{Var}(X_i) = 1$. If $S_n = X_1 + \dots + X_n$ then $\frac{1}{\sqrt{n}}S_n \xrightarrow{n \rightarrow \infty} Z \sim \mathcal{N}(0, 1)$ weakly. In particular $\mathbb{P}(\frac{1}{\sqrt{n}}S_n \leq x) \xrightarrow{n \rightarrow \infty} \mathbb{P}(Z \leq x)$ for all $x \in \mathbb{R}$.*

Proof. For $X = X_1$, the characteristic function $\phi(u) = \phi_X(u) = \mathbb{E}e^{iuX}$ satisfies $\phi(0) = 1$, and $\phi'(u) = i\mathbb{E}Xe^{iuX}$, $\phi''(u) = -\mathbb{E}X^2e^{iuX}$. Since $\phi'(0) = 0$ and $\phi''(0) = -\mathbb{E}X^2 = -1$. Taylor's theorem implies

$$\phi(v) = 1 - \frac{v^2}{2} + o(v^2) \text{ as } v \rightarrow 0$$

Now if $\phi_n(u) = \phi_{\frac{1}{\sqrt{n}}S_n}(u)$, we have

$$\phi_n(u) = \mathbb{E}e^{i\frac{u}{\sqrt{n}}(X_1 + \dots + X_n)} = \prod_{j=1}^n \mathbb{E}e^{i\frac{u}{\sqrt{n}}X_j} = \left[\phi\left(\frac{u}{\sqrt{n}}\right) \right]^n = \left[1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right) \right]^n$$

The complex logarithm satisfies $\log(1+z) = z + o(z)$ as $z \rightarrow 0$. Hence

$$\log \phi_n(u) = n \log \left(1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right) \right) \rightarrow -\frac{u^2}{2}$$

Thus $\phi_n(u) \xrightarrow{n \rightarrow \infty} e^{-\frac{u^2}{2}} = \phi_Z(u)$ so by Levy's theorem, the result follows. \square

Remark: this theorem extends to \mathbb{R}^d via the next proposition, using that $X_n \rightarrow X$ weakly in \mathbb{R}^d iff $\langle X_n, v \rangle \rightarrow \langle X, v \rangle$ in \mathbb{R} for all $v \in \mathbb{R}^d$.

Definition. A random variable X in \mathbb{R}^d is *Gaussian* if $\langle X, v \rangle$ are Gaussian on \mathbb{R} for all $v \in \mathbb{R}^d$.

Proposition. Let X be a Gaussian vector in \mathbb{R}^d . Then

1. $AX + b$ for A an $m \times d$ matrix, $b \in \mathbb{R}^m$, is Gaussian in \mathbb{R}^m .
2. $X \in L^2(\mathbb{P})$, and $\mu = \mathbb{E}X$ and $V = \text{Cov}(X_i, X_j)_{i,j=1}^d$ exist and determine the law μ_X .
3. $\phi_X(u) = e^{i\langle \mu, u \rangle - \frac{\langle u, Vu \rangle}{2}}$, $u \in \mathbb{R}^d$.
4. If V is invertible then μ_X has a probability density function

$$f_X(u) = \frac{1}{(2\pi)^{d/2}} (\det V)^{-1/2} \exp \left\{ -\langle x - \mu, V^{-1}(x - \mu) \rangle \right\}$$

5. Subvectors $X_{(1)}, X_{(2)}$ are independent if and only if $\text{Cov}(X_{(1)}, X_{(2)}) = 0$.

Proof. Example sheet. \square

Proposition (Continuous mapping theorem, Slutsky's lemma). Let $X_n \rightarrow X$ weakly in \mathbb{R}^d as $n \rightarrow \infty$. Then

1. If $h : \mathbb{R}^d \rightarrow \mathbb{R}^k$ is continuous, then $h(X_n) \rightarrow h(X)$ weakly in \mathbb{R}^k .
2. If $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$, then $Y_n \rightarrow X$ weakly.
3. If $Y_n \xrightarrow{P} c = \text{const}$ (on Ω), then $(X_n, Y_n) \rightarrow (X, c)$ weakly in $\mathbb{R}^d \times \mathbb{R}^d$ as $n \rightarrow \infty$. In particular, $X_n + Y_n \rightarrow X + c$ weakly, and $X_n Y_n \rightarrow cX$ weakly by (1).

Proof.

1. Since $g \circ h$ is continuous for continuous g this is clear (look at image measure properties).
2. Take $g : \mathbb{R}^d \rightarrow \mathbb{R}$ bounded and Lipschitz continuous, then

$$|\mathbb{E}g(Y_n) - \mathbb{E}g(X)| \leq \underbrace{|\mathbb{E}g(X_n) - \mathbb{E}g(X)|}_{< \varepsilon/3 \text{ for } n \text{ large}} + \mathbb{E}|g(X_n) - g(Y_n)|$$

And the second term is bounded above by

$$\begin{aligned} & \mathbb{E}|g(X_n) - g(Y_n)| \left(1\{|X_n - Y_n| \leq \frac{\varepsilon}{3\|g\|_{\text{Lip}}}\} + 1\{|X_n - Y_n| > \frac{\varepsilon}{3\|g\|_{\text{Lip}}}\} \right) \\ & \leq \|g\|_{\text{Lip}} \frac{\varepsilon}{3\|g\|_{\text{Lip}}} + 2\|g\|_{\infty} \mathbb{P}(|X_n - Y_n| > \frac{\varepsilon}{3\|g\|_{\text{Lip}}}) \\ & \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \text{ for } n \text{ large enough} \end{aligned}$$

3. First note

$$|(X_n, c) - (X_n, Y_n)| = |Y_n - c| \xrightarrow{\mathbb{P}} 0$$

Also for all $g : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ bounded and continuous

$$\mathbb{E}g(X_n, c) \xrightarrow{n \rightarrow \infty} \mathbb{E}g(X, c)$$

since $g(\cdot, c) : \mathbb{R}^d \rightarrow \mathbb{R}$ is bounded and continuous; and since $X_n \rightarrow X$ weakly, it follows $(X_n, c) \rightarrow (X, c)$ weakly and (2) implies $(X_n, Y_n) \rightarrow (X, c)$ weakly in $\mathbb{R}^d \times \mathbb{R}^d$. □

Laws of Large Numbers

If X_1, \dots, X_n are iid with $\mathbb{E}X_1 = 0$ and $\text{Var}(X_1) = \sigma^2 < \infty$, then by Chebyshev's inequality

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i\right| > \varepsilon\right) \leq \frac{1}{n^2 \varepsilon^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \leq \frac{\sigma^2}{n \varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

So $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mathbb{E}X_1$ as $n \rightarrow \infty$ (Weak Law of Large Numbers).

Proposition. Let $(X_n : n \in \mathbb{N})$ be independent (not necessarily iid) random variables such that $\mathbb{E}X_n = \mu$ and $\mathbb{E}[X_n^4] \leq M$ for all $n \in \mathbb{N}$. Then $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{a.s.}} \mu$ as $n \rightarrow \infty$.

Proof. Setting $Y_n = X_n - \mu$, then $\mathbb{E}Y_n^4 \leq 2^4(\mathbb{E}X_n^4 + \mu^4) < \infty$, so we can assume $\mu = 0$. For distinct indices i, j, k, l

$$0 = \mathbb{E}[X_i X_j X_k X_l] = \mathbb{E}[X_i^3 X_j] = \mathbb{E}[X_i^2 X_j X_k]$$

and by Cauchy-Schwarz

$$\mathbb{E}[X_i^2 X_j^2] \leq (\mathbb{E}[X_i^4])^{1/2} (\mathbb{E}[X_j^4])^{1/2} \leq M.$$

Then

$$\mathbb{E} \left[\left(\sum_{i=1}^n X_i \right)^4 \right] = \mathbb{E} \left(\sum_{i=1}^n X_i^4 \right) + 6 \mathbb{E} \left(\sum_{i \neq j} X_i^2 X_j^2 \right) = nM + 3n(n-1)M \leq 3n^2 M.$$

If $S_n = \sum_{i=1}^n X_i$, then

$$\mathbb{E} \left(\sum_{i=1}^n \left(\frac{S_n}{n} \right)^4 \right) \leq \sum_{n=1}^{\infty} \frac{1}{n^4} 3n^2 M < \infty.$$

So by properties of the integral, $\sum_n \left(\frac{S_n}{n} \right)^4 < \infty$ almost surely, so $\left(\frac{S_n}{n} \right)^4$ and $\frac{S_n}{n} \rightarrow 0$ almost surely as terms of a convergent series. \square

Ergodic Theory

Definition. Let (E, \mathcal{E}, μ) be a σ -finite measure space. A measurable transformation $\Theta : E \rightarrow E$ is *measure preserving* (m.p) if $\mu(\Theta^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{E}$. In this case, for $f \in L^1(\mu)$ we have $\int_E f d\mu = \int_E f \circ \Theta d\mu$.

A measurable map $f : E \rightarrow \mathbb{R}$ is called Θ -invariant if $f \circ \Theta = f$. A set $A \in \mathcal{A}$ is *invariant* if $\Theta^{-1}(A) = A$. Then the collection \mathcal{E}_Θ of Θ -invariant sets is a σ -algebra over E , and f is invariant iff f is \mathcal{E}_Θ -measurable.

Definition. A map Θ is called *ergodic* if the Θ invariant sets $A \in \mathcal{E}_\Theta$ satisfy $\mu(A) = 0$ or $\mu(E \setminus A) = 0$.

Fact: if f is Θ -invariant and Θ is ergodic, then f is constant almost everywhere on E (Example sheet).

Example. On $(E, \mathcal{E}) = ((0, 1], \mathcal{B})$ and μ the Lebesgue measure, the maps $\Theta_a(x) = x + a$ (modulo 1); $\Theta(x) = 2x$ (mod 1), are measure preserving and ergodic unless $a \in \mathbb{Q}$ (Example sheet).

Theorem (Birkhoff's Ergodic Theorem). *Let (E, \mathcal{E}, μ) be σ -finite and let $\Theta : E \rightarrow E$ be measure preserving. For $f \in L^1(\mu)$ define $S_0(f) = 0$, $S_n(f) = \sum_{k=0}^{n-1} f \circ \Theta^k$. Then $\frac{S_n(f)}{n} \xrightarrow{n \rightarrow \infty} \bar{f}$ almost everywhere, where $\bar{f} \in L^1$, $\mu(|\bar{f}|) \leq \mu(|f|)$ and \bar{f} is Θ -invariant.*

The proof of Birkhoff's Ergodic Theorem is based on a key lemma:

Lemma (Maximal ergodic lemma). *For $f \in L^1(E, \mathcal{E}, \mu)$, set*

$$S^* = S^*(f) = \sup_{n \geq 0} S_n(f)$$

Then $\int_{\{S^ > 0\}} f d\mu \geq 0$.*

Proof. Define

$$S_n^* = \max_{0 \leq m \leq n} S_m$$

So $S_n^* \uparrow S^*$. Then $S_m \leq S_n^*$ (on E) for all $m \leq n$ and

$$S_{m+1} = S_m \circ \Theta + f \leq S_n^* \circ \Theta + f \quad (\dagger)$$

Now define $A_n = \{S_n^* > 0\}$ ($\uparrow \{S^* > 0\}$). On A_n , $S_n^* = \max_{1 \leq m \leq n} S_m$ (since $S_0 = 0$), so

$$S_n^* = \max_{1 \leq m \leq n} S_m \leq \max_{0 \leq m \leq n} S_{m+1} \leq^\dagger S_n^* \circ \Theta + f \text{ on } A_n$$

Now integrating this inequality

$$\int_{A_n} S_n^* d\mu \leq \int_{A_n} S_n^* \circ \Theta d\mu + \int_{A_n} f d\mu$$

On A_n^c we must have $S_n^* = 0 \leq S_n^* \circ \Theta$ which implies

$$\int_E S_n^* d\mu \leq \underbrace{\int_E S_n^* \circ \Theta d\mu}_{=\int_E S_n^* d\mu} + \int_{A_n} f d\mu$$

Since Θ is measure preserving. So we obtain

$$\int_{A_n} f d\mu \geq 0$$

Since $f 1_{A_n} \rightarrow f 1_{\{S^* > 0\}}$ and $|f 1_{A_n}| \leq |f| \in L^1(\mu)$ we deduce from the dominated convergence theorem that

$$\int_{\{S^* > 0\}} f d\mu = \lim_{n \rightarrow \infty} \int_{A_n} f d\mu \geq 0$$

□

Proof of Birkhoff's Ergodic Theorem. Notice that

$$\limsup_n \frac{S_n(f)}{n} = \limsup_n \frac{S_{n+1}(f) - f}{n+1} \frac{n+1}{n} = \limsup_n \frac{S_n(f) \circ \Theta}{n}$$

and

$$\liminf_n \frac{S_n(f)}{n} = \liminf_n \frac{S_{n+1}(f) - f}{n+1} \frac{n+1}{n} = \liminf_n \frac{S_n(f) \circ \Theta}{n}$$

are invariant functions and hence their inverse images

$$D = D_{a,b} = \left\{ \liminf_n \frac{S_n(f)}{n} < a < b < \limsup_n \frac{S_n(f)}{n} \right\}$$

are measurable and invariant ($\Theta^{-1}(D) = D$). Without loss of generality we can assume $b > 0$. Then take $B \in \mathcal{E}$, $B \subseteq D$ such that $\mu(B) < \infty$ and set $g = f - b1_B \in L^1(\mu)$. Then

$$S_n(g) = S_n(f) - bS_n(1_B) \geq S_n(f) - bn > 0 \text{ on } D \text{ for some } n \in \mathbb{N} \quad (*)$$

Now we apply the maximal ergodic lemma with $E = D$, which is still Θ -invariant since

$$\begin{aligned} \mu|_D(A) &= \mu(A \cap D) = \mu(\Theta^{-1}(A \cap D)) = \mu(\Theta^{-1}(A) \cap \Theta^{-1}(D)) \\ &= \mu(\Theta^{-1}(A) \cap D) = \mu|_D(\Theta^{-1}(A)) \end{aligned}$$

Note that $\{S^*(g)0\} \subseteq D$ as we are restricting and $D \subseteq \{S^*(g) > 0\}$ by (*). Then

$$0 \leq \int_{\{S^* > 0\}} g d\mu = \int_D g d\mu = \int_D f d\mu - b\mu(B)$$

so $b\mu(B) \leq \int_D f d\mu$. By σ -finiteness, this inequality extends to D - take an approximating sequence $B_n \uparrow D$, $\mu(B_n) < \infty$ and take limits, so $b\mu(D) = b \lim_n \mu(B_n) \leq \int_D f d\mu$. Repeating this argument with $-f, -a$ we deduce

$$(-a)\mu(D) \leq \int_D (-f) d\mu$$

which combined gives

$$b\mu(D) \leq \int f d\mu \leq a\mu(D)$$

but $a < b$, so necessarily $\mu(D) = 0$.

Now define

$$\begin{aligned} \Delta &= \left\{ \liminf_n \frac{S_n}{n} < \limsup_n \frac{S_n}{n} \right\} \\ &= \bigcup_{\substack{a,b \in \mathbb{Q} \\ a < b}} D(a,b) \end{aligned}$$

hence since $\mu(D(a,b)) = 0$ by the previous, we see $\mu(\Delta) \leq \sum_{a < b} \mu(D(a,b)) = 0$. On Δ^c , $\frac{S_n}{n}$ converges in $[-\infty, \infty]$ and we define the invariant function

$$\bar{f} = \begin{cases} \lim_n \frac{S_n}{n} & \text{on } \Delta^c \\ 0 & \text{on } \Delta \end{cases}$$

so that $\frac{S_n}{n} \rightarrow \bar{f}$ almost everywhere as $n \rightarrow \infty$. Finally, since $\mu(|f \circ \Theta^n|) = \mu(|f|)$ we have $\mu(|S_n|) \leq n\mu(|f|)$ and thus (by Fatou's lemma)

$$\mu(|\bar{f}|) = \mu\left(\liminf_n \left|\frac{S_n}{n}\right|\right) \leq \liminf_n \mu\left(\left|\frac{S_n}{n}\right|\right) \leq \mu(|f|) < \infty$$

So in particular $|\bar{f}| < \infty$ almost everywhere. □

Consider $E = \mathbb{R}^{\mathbb{N}} = \{x = (x_n \in \mathbb{R} : n \in \mathbb{N})\}$. Consider ‘rectangles’

$$\mathcal{C} = \{A = \times_{n=1}^{\infty} A_n, A_n \in \mathcal{B}, A_n = \mathbb{R} \ \forall n > N, \text{ some } N \in \mathbb{N}\}$$

which form a π -system which generate the *cylindrical* σ -algebra $\sigma(\mathcal{C})$. One shows that $\sigma(\mathcal{C}) = \sigma(f_n : n \in \mathbb{N})$ where $f_n(x) = x_n$ (coordinate projection on E) and also $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ for the product topology.

Let $(X_n : n \in \mathbb{N})$ be a sequence of iid random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with marginal distributions $\mu_{X_n} = m$ for all n (this exists by a theorem from earlier). Define a map $X : \Omega \rightarrow E$ such that $X(\omega) = (X_1(\omega), X_2(\omega), \dots)$. X is \mathcal{F} - $\sigma(\mathcal{C})$ measurable since for all $A \in \mathcal{C}$,

$$\begin{aligned} X^{-1}(A) &= \{\omega : X_1(\omega) \in A_1, \dots, X_N(\omega) \in A_N\} \\ &= \bigcap_{n \leq N} X_n^{-1}(A_n) \in \mathcal{F} \end{aligned}$$

Denote by μ the image measure $\mu = \mathbb{P} \circ X^{-1}$ on $\mathbb{R}^{\mathbb{N}}$, which is the unique product probability measure on $\mathbb{R}^{\mathbb{N}}$ such that

$$\begin{aligned} \mu(\times_{n=1}^{\infty} A_n) &= \lim_{N \rightarrow \infty} \mu(\times_{n=1}^N A_n) = \lim_{N \rightarrow \infty} \mathbb{P}(X_1 \in A_1) \dots \mathbb{P}(X_N \in A_N) \\ &= \lim_{N \rightarrow \infty} \prod_{i=1}^N m(A_i) \\ &= \prod_{n=1}^{\infty} m(A_n) \end{aligned}$$

We call $(E, \mathcal{E}, \mu) = (\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}), m^{\mathbb{N}})$ the canonical model for an infinite sequence of random variables of law m .

On E , define the shift map $\Theta : E \rightarrow E$ by $\Theta((x_1, \dots, x_n)) = (x_2, x_3, \dots)$.

Theorem. *The shift map is measure preserving and ergodic.*

Proof. For rectangles $A \in \mathcal{C}$,

$$\begin{aligned} \mu(A) &= \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) \\ &= \mathbb{P}(X_1 \in A_1) \dots \mathbb{P}(X_N \in A_N) \\ &= \prod_{n=1}^N m(A_n) \\ &= \mathbb{P}(X_2 \in A_1) \dots \mathbb{P}(X_{N+1} \in A_N) \\ &= \mu(x : \Theta(x) \in A) = \mu(\Theta^{-1}(A)) \end{aligned}$$

so Θ is measure preserving. To show Θ is ergodic, recall the tail- σ -algebra $\mathcal{T} = \bigcap_n \mathcal{T}_n$, where $\mathcal{T}_n = \sigma(X_k : k \geq n+1)$. Clearly for $A \in \sigma(\mathcal{C})$ then

$$\Theta^{-n}(A) = \{x \in \mathbb{R}^{\mathbb{N}} : (x_{n+1}, \dots) \in A\} \subseteq \mathcal{T}_n$$

Now if A is invariant, $A = \Theta^{-n}(A) \in \mathcal{T}_n$ for all n , so $A \in \mathcal{T}$ and by Kolmogorov's 0-1 law, $\mu(A) = 1$ or $\mu(A) = 0$. \square

By Birkhoff's ergodic theorem, if $f \in L^1(\mu)$, then

$$\frac{S_n(f)}{n} = \frac{f + f \circ \Theta + \dots + f \circ \Theta^{n-1}}{n} \xrightarrow{n \rightarrow \infty} \bar{f} \text{ almost surely}$$

with $\bar{f} = \text{constant}$ almost surely. By von-Neumann's L^p -ergodic theorem, convergence holds in fact in L^1 (we will prove this shortly).

Theorem. Assume $\int_{\mathbb{R}} |x| d\mu(x) < \infty$ and set $\int_{\mathbb{R}} x d\mu(x) = \nu$. Then

$$\mu \left(x \in \mathbb{R}^{\mathbb{N}} : \frac{x_1 + x_2 + \dots}{n} \rightarrow \nu \right) = 1.$$

Proof. Let $f(x) = x_1$, which is in $L^1(\mu)$ since $\int_E |f| d\mu = \int_{\mathbb{R}} |x| d\mu(x) < \infty$ so by the ergodic theorem

$$\mu \left(\frac{x_1 + x_2 + \dots + x_n}{n} \xrightarrow{n \rightarrow \infty} \nu \right) = \mu \left(\frac{S_n(f)}{n} \rightarrow \bar{f} \right) = 1$$

where we also use von-Neumann's theorem to deduce

$$\bar{f} = \mu(\bar{f}) = \lim_n \mu \left(\frac{S_n(f)}{n} \right) = \lim_n \frac{n}{n} \nu = \nu$$

(\bar{f} is constant almost surely). \square

Theorem (Strong law of large numbers). Let $(X_n : n \in \mathbb{N})$ be iid random variables such that $\mathbb{E}|X_1| < \infty$. Then $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} \mathbb{E}X_1$ almost surely.

Proof. Inject $X : \Omega \rightarrow E = \mathbb{R}^{\mathbb{N}}$ as before and notice

$$\mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} \mathbb{E}X \right) = \mu \left(x : \frac{x_1 + \dots + x_n}{n} \xrightarrow{n \rightarrow \infty} \nu \right) = 1$$

\square

Remarks:

1. The hypothesis $\mathbb{E}|X_1| < \infty$ is necessary (Example sheet).
2. The law of the iterated logarithm (LIL) says

$$\limsup_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{\sqrt{2n \log \log n}} = 1 \text{ almost surely}$$

$$\liminf_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{\sqrt{2n \log \log n}} = -1 \text{ almost surely}$$

so the Central Limit Theorem does not hold almost surely.

Theorem (von Neumann's ergodic theorem). In the setting of Birkhoff's theorem, assume further that $\mu(E) < \infty$ and $f \in L^p(E)$, for some $1 \leq p < \infty$. Then

$$\frac{S_n(f)}{n} \xrightarrow{n \rightarrow \infty} \bar{f} \text{ in } L^p$$