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 $\mu : \mathcal{E} \to [0, \infty]$ 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

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

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

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

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 $B_n \in \mathcal{E}$ $\underline{\text{disjoint}}$ such that $\bigcup_n A_n = \bigcup_n B_n$. Indeed, define $\tilde{A}_n = \bigcup_{j \le n} A_j$, 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 $\mu : \mathcal{A} \to [0, \infty]$ such that $\mu(\emptyset) = 0$. We say μ is

1. *increasing* if $\mu(A) \le \mu(B)$ 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 $\mu(\cup_n A_n) = \sum_n \mu(A_n)$ for any $(A_n : n \in \mathbb{N})$ where $A_n \in \mathcal{A}$ disjoint and $\cup_n A_n \in \mathcal{A}$.
4. *countably sub-additive* if $\mu(\cup_n A_n) \le \sum_n \mu(A_n)$ for all $(A_n : n \in \mathbb{N})$ such that $\cup_n A_n \in \mathcal{A}$

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

$$\mu^*(B) \leq \sum_n \mu^*(B_n) \tag{†}$$

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 (†) follows since ε was arbitrary.

Step 2: μ^* extends μ. Let $A \in \mathcal{A}$. Clearly $A = A \cup \emptyset \cup \ldots \cup \emptyset$, so by definition of μ^*, $\mu^*(A) \leq \mu(A) + 0 + \ldots + 0$. 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

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

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 $B \subseteq (B \cap A) \cup (B \cap A^c) \cup \emptyset \ldots \cup \emptyset = \bigcup_n B_n$ so by (†) we have

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

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

$$\mu^*(B) + \varepsilon \geq \sum_n \mu(A_n) \tag{$\circ$}$$

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 (†) we clearly have

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

and

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

so we only need two converse inequalities. Similar to before

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

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

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

and by (†)

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

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

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

$\square$

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 conatins $\mathcal{A}$ also contains $\sigma(\mathcal{A})$.*

Proof. Define

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

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) $\mu_i(B \setminus B_n) \to \mu_i(\emptyset) = 0$ for $i = 1, 2$. This implies for $\mu_i(B) < \infty$ that $\mu_i(B_n) \to \mu_i(B)$ as $n \to \infty$ for both $i = 1, 2$. But then

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

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

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

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

We will do $d = 1$ first.

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

$$\mu((a, b]) = b - a,\ \forall a < b \tag{†}$$

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

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

(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 $D_k = \bigcup_j (D_K \cap C_k)$, so

$$\mu(A) = \mu\left(\bigcup_j C_j\right) = \sum_j \mu(C_j) = \sum_j \mu\left(\bigcup_k (C_j \cap D_k)\right)$$

$$= \sum_{j,k} \mu(C_j \cap D_k) = \ldots = \mu\left(\bigcup_k D_k\right) = \mu(A)$$

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 $\mu(A_n) \to 0$.

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

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

Hence since μ is increasing

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

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

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

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

$$K_N = \overline{C_1} \cap \ldots \cap \overline{C}_N$$

($\overline{C}_i$ 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 (†) must exist.

For uniqueness, suppose μ, λ measures such that (†) 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

$$\mu(A) = \mu\left(\bigcup_n A \cap (n, n+1]\right) = \sum_n \mu(A \cap (n, n+1]) = \sum_n \mu_n(A)$$

$$= \sum_n \lambda_n(A) = \ldots = \lambda(A)$$

so $\lambda = \mu$. □

Remarks:

1. a set $B \in \mathcal{B}$ is called a Lebesgue null set if $\mu(B) = 0$. Can write $\{x\} = \bigcap_n \left(x - \frac{1}{n}, x\right]$ 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 exted to $\mathcal{P}(E)$,

Theorem (Banach,Kuretowski)**.** *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 $(E, \mathcal{E}, \mu)$ (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 $\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n)$ [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.

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

□

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 $1 - a \leq e^{-a}$,

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

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

$$\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} A_m\right) = 1 - \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} A_m^c\right)$$

$$\geq 1 - \sum_n \underbrace{\mathbb{P}\left(\bigcap_{m \geq n} A_m^c\right)}_{=0} = 1$$

$\square$

2 Measurable functions

Let $(E, \mathcal{E})$, $(G, \mathcal{G})$ be measurable spaces and let $f : E \to G$. 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 $f : (E, \mathcal{E}) \to \mathbb{R}$ 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: $f^{-1}\left(\bigcup_i A_i\right) = \bigcup_i f^{-1}(A_i)$ 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 $f^{-1}((-\infty, y]) = \{x \in E : f(x) \le y\} \in \mathcal{E}$ for all $y \in \mathbb{R}$. Moreover, if E is a topological space with $\mathcal{E} = \mathcal{B}(E)$, then if $f : E \to \mathbb{R}$ 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 $\{f_i : E \to (G, \mathcal{G}), i \in I\}$ we can make them all measurable for

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

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 $0 \le f_n \uparrow f$ pointwise on E, then $f \in \mathcal{V}$.*

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

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 Dynkind'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 $f : E \to \mathbb{R}$ 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 $j = 0, \ldots, n2^n - 1$, 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 $|f_n - f| \leq 2^{-n} \to 0$. 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 $f : (E, \mathcal{E}) \to (G, \mathcal{G})$ 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

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

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

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

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

Then f is monotone increasing, left-continuous and

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

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 $f(x_n) \to f(x)$ as $x_n \to x$. □

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 $\mathcal{B}$-$\mathcal{B}(I)$ 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 $g(y_n) \to g(y)$ 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. □

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 $\mathcal{F}$-$\mathcal{E}$ measuable map

$$X : \Omega \to E$$

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. $\lim_{z\to-\infty} F_X(z) = \mu_X(\emptyset) = 0$ and $\lim_{z\to\infty} F_X(z) = \mu_X(\mathbb{R}) = \mathbb{P}(\Omega) = 1$

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 obatin a random variable $X : \Omega \to \mathbb{R}$ by

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

with distribution function F_X.

Definition. A countable collection $(X_i : (\Omega, \mathcal{F}, \mathbb{P} \to (E, \mathcal{E}))$ 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

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

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 $\omega = \sum_{i=1}^{n} \omega_i 2^{-i}$, 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 $(x_i)_{i=1}^{n} \in \{0,1\}^n$

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

So the R_n are all independent. Now take a bijection $m : \mathbb{N}^2 \to \mathbb{N}$ 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 $\left(\frac{i}{2^m}, \frac{i+1}{2^m}\right]$, $i = 0, \dots, 2^m - 1$, $m \in \mathbb{N}$, 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 sequene 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 $f_n \to f$ μ-*almost always* if

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

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

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

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

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

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

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

Proof.

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

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

so $\liminf_n \mu(|f_n| \le \varepsilon) \ge \mu(E)$. So we see $\limsup_n \mu(|f_n| > \varepsilon) \le \mu(E) - \mu(E) = 0$, so $\mu(|f_n| > \varepsilon) \to 0$ as $n \to \infty$ 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 $\mu(|f_{n_k}| > \frac{1}{k}) \le \frac{1}{k^2}$ so

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

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

$\square$

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 $1_{A_n} \to 0$ in $\mathbb{P}$-probability since $\mathbb{P}(1_{A_n} > \varepsilon) = \mathbb{P}(A_n) = \frac{1}{n} \to 0$ 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 $1_{A_n} \not\to 0$ 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

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

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 $\mathcal{T} = \sigma(X_{n+1}, X_{n+2}, \ldots)$ 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 $Y : (\Omega, \mathcal{T}) \to (\mathbb{R}, \mathcal{B})$ is meaurable, then Y is constant almost surely.*

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

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

and note that the π-system of sets

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

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 $\mathcal{F}_\infty = \sigma(X_1, X_2, \ldots)$, then $\bigcup_n \mathcal{F}_n$ is a π-system generating $\mathcal{F}_\infty$, 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 $\mathcal{F}_\infty$ 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 $f : (E, \mathcal{E}, \mu) \to \mathbb{R}$ measurable or "integrable" we will define the integral with respect to μ:

$$\mu(f) = \int_E f \mathrm{d}\mu = \int_E f(x) \mathrm{d}\mu(x)$$

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

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

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

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

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

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

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

For $f : (E, \mathcal{E}, \mu) \to \mathbb{R}$ 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 $f_n, f : (E, \mathcal{E}, \mu) \to \mathbb{R}$ 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 $f_n(x) \to f(x)$ for all $x \in E$). Then $\mu(f_n) \to \mu(f)$ as $n \to \infty$.*

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, $[\tilde{f}_n]_n = \bar{f}_n = \min(2^{-n}\lfloor 2^n f_n \rfloor, n)$. 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

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

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 $\mu(f) = \mu(\lim_n f_n) = \lim_n \mu(f)$, so we can interchange $\int(\cdot)\mathrm{d}\mu$ and the limit. If $g_n \geq 0$, then $\mu\left(\sum_n g_n\right) = \sum_n \mu(g_n)$. 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 $f, g : (E, \mathcal{E}, \mu) \to \mathbb{R}$ 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. □

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

Theorem. *Let $f, g : (E, \mathcal{E}, \mu) \to \mathbb{R}$ 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 $\mu(f) = \mu(f^+) - \mu(f^-) = 0 - 0$. □