Probability and Measure

November 24, 2017

Contents

Check James Norris notes; Terrence Tao; *An introduction to measure theory*

1 Measure Theory

1.1 History

What is a measure? It is a function $\mathbb{R}^d \to \mathbb{R}$ defined on subsets, which we would like it to satisfy the following properties:
- Non-negativity: $m(E) \geq 0$ for all E;
- Empty set: $m(\phi) = 0$;
- Additivity: $m(E \cup F) = m(E) + m(F)$ for any two disjoint sets E and F; • Normalization: $m([0,1]^d) = 1$;
- Translation invariant: $m(E + x) = m(E)$ for all E and all $x \in \mathbb{R}^d$.

It is possible to construct pathological 'measures' satisfying those axioms and defined *on all* subsets of $\mathbb{R}^d$, but they won't be 'nice'. When mathematicians construct such measures, they usually do so on a restricted class of subsets. Otherwise this may lead to contradictions (see below). In fact, if $d \geq 3$, we can show that there's no measure $m : \mathbb{R}^d \to [0, \infty]$ which is also rotation invariant, i.e. $m(gE) = m(E)$ for any Euclidean rotation g in $\mathbb{R}^d$.

Now consider Jordan Measure. First we define the measure of a box $B \subseteq \mathbb{R}^d$ with edges parallel to the axes to be the product of the lengths of its sides.

Definition. An elementary subset of $\mathbb{R}^d$ is a finite union of boxes.

Remark. Every elementary set can be written as a finite union of disjoint boxes. The family of elementary sets is stable under finite unions, finite intersections, and difference. Also, if we can write a set E as disjoint union of boxes in two different ways, then the sum of their measures coincide. As a result, for $E = \cup B_i$ for disjoint B_i, it makes sense to define $m(E)$ as $\sum m(B_i)$.

Definition. A subset E of $\mathbb{R}^d$ is *Jordan measurable* if $\forall \varepsilon > 0$, there exists elementary sets A, B s.t. $A \subseteq E \subseteq B$ and $m(B \setminus A) < \varepsilon$.

Note that if E is Jordan measurable, then

$$\inf\{m(B)|E \subseteq B, B \text{ elementary}\} = \sup\{m(A)|A \subseteq E, A \text{ elementary}\}$$

We then define the Jordan measure of E as this limit. It's an exercise to check that m satisfies all the axioms spelled out earlier.

We can draw a connection between Jordan measure and Riemann integral. Recall the definition of Riemann integrability: $f : [a, b] \to \mathbb{R}$, $a < b$ is Riemann integrable if all its Riemann sums converge, i.e. $\exists I = I(f) \in \mathbb{R} s.t. \forall \varepsilon > 0 \exists \delta > 0 \forall$ partition P of $[a, b]$ of width $< \delta$, $|S(f, P) - I(f)| < \varepsilon$, where $S(f, P)$ is the Riemann sum of f by P.

Now we draw the link:

Proposition. f is Riemann integrable iff $E^+ = \{(x, t) \in \mathbb{R}^2, 0 \leq t \leq f(x)\}$ and $E^- = \{(x, t) \in \mathbb{R}^2, f(x) \leq t \leq 0\}$ are Jordan measurable.

1.2 Lebesgue Measure

Example. Consider the function $1_{\mathbb{Q}}(x)$, which is 1 if x is rational and 0 otherwise. This is not Riemann integrable, and $\mathbb{Q}$ or even $\mathbb{Q} \cap [0,1]$ is not a Jordan measurable subset of $\mathbb{R}$.

In fact, no dense countable subset of an interval in $\mathbb{R}$ is Jordan measurable.

There is a problem if we try to define measure with limits of functions, because a pointwise limit of Riemann integrable functions is not always Riemann integrable. One such example is the set of functions $f_n = 1_{\frac{1}{n!}\mathbb{Z} \cap [0,1]}$, which is Riemann integrable, but as we've seen, the limit as $n \to \infty$, $1_{\mathbb{Q}}(x)$, is not.

Lebesgue's idea is to remove the containment of A by E, as well as to allow countable union of boxes to be elementary sets as well. We then reach the following definition:

Definition. A set $E \subseteq \mathbb{R}^d$ is called *Lebesgue measurable* if $\forall \varepsilon > 0$, there exists a countable family of boxes $(B_i)_{i \geq 1}$ such that $E \subseteq \bigcup_{i \geq 1} B_i$, and $m^*(\bigcup B_i \setminus E) < \varepsilon$. Here m^* is the *Lebesgue outer measure*, defined by: if $F \subset \mathbb{R}^d$, then

$$m^*(F) = \inf \left\{ \sum_{n \geq 1} m(B_n), F \subseteq \bigcup_{n \geq 1} B_n \right\}$$

where B_n are boxes. Note that here we allow countably many B_n instead of just finitely many.

Note that m^* is defined for all subsets of $\mathbb{R}^d$ (least upper bound property of $\mathbb{R}$), but m^* is not additive on all subsets of $\mathbb{R}^d$. If we let

$$m^{*,J}(F) = \inf \left\{ \sum_i^N m(B_i), F \subseteq \bigcup_1^N B_i \right\}$$

where B_i are boxes (this is called the Jordan outer measure), then obviously $m^*(F) \leq m^{*,J}(F)$. The important part is that sometimes $m^*(F) < m^{*,J}(F)$: consider $F = \mathbb{Q} \cap [0,1]$. If we want to cover F by finitely many intervals, we have to cover the whole $[0,1]$, so $m^{*,J}(F) = 1$. However if we are allowed a countable family of boxes, then we have $m^*(F) = 0$ (not very obvious – why?).

As an exercise, check that m^* satisfies the following properties: $m^*(\phi) = 0$, $m^*(E) \leq m^*(F)$ if $E \subseteq F$ (monotone), and $m^*(\bigcup_{i \geq 1} E_i) \leq \sum_{i \geq 1} m^*(E_i)$ (countable subadditivity), and m^* is *translation invariant*.

Now we give an example to show why m^* is not additive on all subsets of $\mathbb{R}^d$.

Example. (Vitali's example)
Let E be a set of representatives of the cosets of the subgroup $(\mathbb{Q}, +) \subseteq (\mathbb{R}, +)$. WLOG we choose $E \subseteq [0,1]$ (to be clear, this means that $\forall x \in \mathbb{R} \exists! e \in E$ s.t. $x - e \in \mathbb{Q}$. So the family of sets $\{E + \omega\}$ for all $\omega \in \mathbb{Q}$ is a disjoint family of subsets of $\mathbb{R}$ (a partition).

By translation invariance, $m^*(E + r) = m^*(E)$ for all $r \in \mathbb{Q}$.

For any distinct $\omega_1, ..., \omega_N \in \mathbb{Q} \cap [0, 1]$, if m^* were additive, then we have

$$m^*(\bigcup_{i=1}^{N} \omega_i + E) = Nm^*(E)$$

but

$$\bigcup_{1}^{N} \omega_i + E \subseteq [0, 2]$$

so we must have

$$m^*(\bigcup \omega_i + E) \leq m^*([0, 2]) \leq 2$$

But N is arbitrary. so we must have $m^*(E) = 0$. Lastly, since

$$[0, 1] \subseteq \bigcup_{\omega \in \mathbb{Q}} E + \omega$$

by countable subadditivity and translation invariance we get

$$m^*([0, 1]) \leq \sum_{r \in \mathbb{Q}} m^*(E + r) = 0$$

So $1 \leq 0$. Contradiction.

Remark. It is worth notice that this example requires the use of the axiom of choice to construct E as there is no apparent way to choose the elements of E explicitly. So a better version of conclusion of the above example should be that AC is not compatible with additivity of m^*.

Next time we will define the Lebesgue measure of a Lebesgue measurable set E as $m(E) := m^*(E)$, and we will show that m is additive on Lebesgue measurable sets. We will also show that m^* is indeed a measure.

Now lets consider the *middle thirds Cantor set*. It is a compact subset C of $[0, 1]$, defined by the following way: we start with $[0, 1]$, and cut it into three equal disjoint parts, and remove the central part. We do the same recursively for each of the remaining parts, and in the limit we get a subset C of $[0, 1]$ and we call it the Cantor set. Also we denote $I_0 = [0, 1]$, $I_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, and so on.

Note that we can express this 'infinite' process in another way: if we write the real numbers in $[0, 1]$ in base-3 decimal expansion, then C contains exactly those that do not have any 1 on any digit of their decimal part.

Now I_n is a finite union of intervals, so it is an elementary set. We see that

$$m(I_n) = 2^n \cdot \frac{1}{3^n}$$

so $\lim_{n \to \infty} m(I_n) = 0$, i.e. C is Jordan measurable with measure 0.

We know every Jordan measurable set is Lebesgue measurable (clear from definition).

Aside: what is the outer measure of a Vitali set? Let E be the Vitali set, i.e. the set of representatives of cosets of $(\mathbb{Q}, +)$ in $(\mathbb{R}, +, E \subseteq [0, 1]$.

We know $m^*(E) > 0$ by the argument given above, but we could choose $E \subseteq [0, \frac{1}{n}]$ or $E \subseteq$ any interval. The answer is it depends on the choice of E, it must be $m^*(E) > 0$ can be arbitrarily small.

Recall last time where we defined the outer measure of an arbitrary subset $E \subseteq \mathbb{R}^d$,

$$m^*(E) = \inf \left\{ \sum_{n \geq 1} m(B_n), E \subseteq \bigcup_n B_n \right\}$$

for B_i boxes in $\mathbb{R}^d$. It doesn't matter in this definition if we use close boxes or open boxes (because of inf). We also defined Lebesgue measurable last time. Again we can take open boxes or close boxes for that definition as well.

Definition. A *null* set $E \subset \mathbb{R}^d$ is a subset E s.t. $m^*(E) = 0$.

Remark. Every null set is Lebesgue measurable. (not obvious)

Recall that the outer measure m^* satisfies $m^*(\phi) = 0$, $m^*(E) \leq m*(F)$ if $E \subseteq F$, and $m^*(\bigcup_n E_n) \leq \sum_n m^*(E_n)$.

Proposition. (1)
The family of Lebesgue measurable subsets of $\mathbb{R}^d$, $\mathcal{L}$, is stable under (a) countable unions, (b) complementation, and (c) countable intersections.

Proposition. (2) Every closed or open subset of $\mathbb{R}^d$ is Lebesgue measurable.

Proof. (of (1))
Obviously (c) follows from (a) and (b) because complement of intersection is the union of complements.

We prove (a) first. Let $(E_n)_{n \geq 1}$ be a countable family in $\mathcal{L}$. Pick $\varepsilon > 0$. By definition, $\exists C_n := \cup_{i \geq 1} B_i^{(n)}$ where $B_i^{(n)}$ are boxes, s.t. $m^*(C_n \setminus E_n) < \frac{\varepsilon}{2^n}$.

Now $\cup E_n \subset \cup C_n = \cup_{n,i} B_i^{(n)}$ is still a countable union of boxes, and

$$m^*\left(\bigcup B_i^{(n)} \setminus \bigcup E_n \right) \leq \sum_n m^*(C_m \setminus E_n)$$
$$\leq \sum_{n \geq 1} \frac{\varepsilon}{2^n} \leq \varepsilon$$

This proves (a).

Now let's prove (b). We start with $E \subset \mathcal{L}$. By definition, $\forall n, \exists C_n$ a countable union of boxes s.t. $E \subset C_n$ and $m^*(C_n \setminus E) \leq \frac{1}{n}$. Now $E \subseteq C_n$. Taking complements we get $C_n^c \subseteq E^c$ and $C_n \setminus E = E^c \setminus C_n^c$.

We could take C_n to be a union of open boxes. Then C_n is open and C_n^c is closed. Then by part (a), C_n^c is closed. But

$$\bigcup C_n^c \subseteq E^c$$

and

$$m^*(E^c \setminus C_n^c) \leq m^*(E^c \setminus C_n^c) = m^*(C_n \setminus E) < \frac{1}{n}$$

Hence $m^*(E^c \setminus \cup_n C_n^c) = 0$. So

$$E^c = \cup_n C_n^c \cup E^c \setminus \cup_n C_n^c$$

but both parts of the union are in $\mathcal{L}$. So by (a), $E^c \in \mathcal{L}$. This shows (b) by using prop (2). $\qquad\square$

Proof. (of (2))
We first prove a lemma: Every open subset of $\mathbb{R}^d$ is a countable union of open boxes (i.e. $\mathbb{R}^d$ has a countable base). To show this we just consider all the boxes centred at rational points and with rational side lengths, which is a countable base.

An immediate consequence is that open sets are measurable by prop (1a). Note that we can't use prop (1b) to prove the same for closed subsets because we used prop (2) in the proof of (1b).

To show that closed subsets are also Lebesgue measurable, it's enough to show that compact subsets of $\mathbb{R}^d$ are Lebesgue measurable because compact is equivalent to closed and bounded in $\mathbb{R}^d$, and every closed subset $F \in \mathbb{R}^d$ is a countable union of compact subsets (to see this, consider the closed annuli centred at origin with radius integers).

Now let $F \subseteq \mathbb{R}^d$ be a compact subset. By definition of $m^*(F)$, for every $k \geq 1$, there exists a countable union of boxes $\bigcup B_n^{(k)}$ such that $F \subseteq \bigcup B_n^{(k)}$ and $m^*(F) + \frac{1}{2}k \geq \sum_n m(B_n^{(k)})$ (k is just ε here). Note that up to subdividing each $B_n^{(k)}$ into a finite number of smaller boxes, we can assume that each $B_n^{(k)}$ has diameter $\leq \frac{1}{2^k}$. WLOG we assume that each boxes meets F (otherwise we can just remove that box). This is then an open cover of F. Since F is compact, there is a finite subcover, i.e. we can pick finitely many boxes from $B_n^{(k)}$ whose union still covers F.

Now let $U_k = \bigcup_{n=1}^{N_k} B_n^{(k)}$, where $F \subseteq U_k$ for all k, and F meets each box $B_n^{(k)}$. Then we must have $F = \bigcup_{k \geq 1} U_k$, because if $x \in \cup U_k$, $\forall k \exists x_k \in F$, that x and x_k lie in the same box $B_n^{(k)}$, so $||x - x_k||_\infty \leq \frac{1}{2^k}$ i.e. x is a limit point of F; then by compactness of F we know x is in F.

We need to show that $M^*(U_k \setminus F)$ tends to 0 and we're done. We claim that if A, B are two disjoint compact subsets of $\mathbb{R}^d$, then m^* is additive on them. To prove that, we choose disjoint covers for them by open boxes (think of they are separated by some positive distance due to compactness – see MT in IB).

We apply above to $A = F$, $B = \bar{U}_k \setminus U_{k+1}$, and get

$$
\begin{aligned}
m^*(F) + m^*(\bar{U}_k \setminus U_{k+1}) &\leq m^*(\bar{U}_k \setminus U_{k+1} \cup F) \\
&\leq m^*(\bar{U}_k) \\
&= m^*(U_k) \\
&\leq m^*(F) + \frac{1}{2^k}
\end{aligned}
$$

So

$$
m^*(U_{k+1} \setminus U_k) \leq \frac{1}{2^k}
$$

Since $F = \bigcup_k U_k$, by countable subadditivity of m^*, we get

$$
m^*(U_k \setminus F) \leq \sum_{i \geq k} m^*(U_i \setminus U_{i+1}) \leq \sum_{i \geq k} \frac{1}{2^i} \leq \frac{1}{2^{k-1}} \to 0
$$

as $k \to \infty$.

$\square$

Definition. Let X be a set. A family $\mathcal{A}$ of subsets of X is a *σ-algebra* if it contains the empty set and is closed under countable union and complementation.

The σ here in the name stands for 'countable' in some sense. σ-algebras are a special type of *Boolean algebras*. The difference between Boolean algebra and sigma algebra is that Boolean algebra allows countable unions in its definition.

Example. • Let $B = \mathcal{P}(X)$, the power set of X. Then B is a Boolean algebra.
• $B = \{\phi, X\}$ is a trivial Boolean algebra.
• Take a partition of X into finitely many pieces $X = \cup_{i=1}^n P_i$. Take $B := \{A \subset X \mid \exists I \subset \{1, ..., \mathbb{N}\}, A = \cup_{i \in I} P_i\}$. Then B is also a (finite) Boolean algebra.

It should be quite obvious that every finite Boolean algebra is of the above form.

Definition. A *measurable space* is a couple $(X, \mathcal{A})$ where X is a set and $\mathcal{A}$ is a σ-algebra of subsets of X.

Definition. A *measure* on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, \infty]$ s.t. $\mu(\phi) = 0$ and μ is countably additive.

Remark. Sometimes people call a set function a function from a Family of subsetes of X to $[0, \infty]$ s.t. $\mu(\phi) = 0$.

Definition. If X is a set, $\mathcal{A}$ is a σ-algebra on X, μ is a measure on $\mathcal{A}$, then $(X, \mathcal{A}, \mu)$ is called a *measure space*.

Last time we've shown that $(\mathbb{R}^d, \mathcal{L})$ is a measurable space.

Definition. (Lebesgue measure)
If $E \in \mathcal{L}$ (the set of Lebesgue measurable sets), we define the Lebesgue measure of E as

$$
m(E) = m^*(E)
$$

the outer measure of E which we've previously defined. Recall that

$$m^*(E) = \inf \left\{ \sum_{n \geq 1} m(B_n), E \subset \bigcup_n B_n \right\}$$

for B_n boxes.

Proposition. m is countably additive on $\mathcal{L}$. Hence, $(\mathbb{R}^d, \mathcal{L}, m)$ is a measure space.

Proof. A special case is 2 disjoint compact subsets $A, B \subseteq \mathbb{R}^d$. We've seen that (check) $m^*(A) + m^*(B) = m^*(A \cup B)$. Suppose now $\{E_n\}_n$ is a family of pairwise disjoint compact subsets of $\mathbb{R}^d$. Then

$$m^*(\cup_1^N E_i) = \sum_1^N m^*(E_i)$$

by iterating the special case. So

$$\sum_{i=1}^N m^*(E_i) \leq m^*(\cup_1^\infty) E_i \leq \sum_{i \geq 1} m^*(E_i)$$

by monotonicty of m^* and countable subadditivity respectively. Now let N tend to ∞ and we get

$$\sum_{i=1}^\infty m^*(E_i) = m^*(\cup_1^\infty E_i)$$

Now let's only assume that the E_n's are bounded sets.

Lemma. If E is a Lebesgue measurable subset of $\mathbb{R}^d$, then $\forall \varepsilon > 0$, there exists $U \subseteq \mathbb{R}^d$ an open set with $E \subseteq U$ and $m^*(U \setminus E) < \varepsilon$. Also, $\forall \varepsilon > 0$, there exists $F \subseteq E$ closed set with $m^*(E \setminus F) < \varepsilon$.

Proof. Note that the second assertion follows from the fist one applied to E^c. So let's prove the first one – but this is just the definition of E being Lebesgue measurable: $\forall \varepsilon \exists \cup B_n$s.t. $E \subseteq \cup B_n$ and $m^*(\cup B_n \setminus E) < \varepsilon$.

Recall we could take the B_n to be open boxes, then we just set $U = \cup_n B_n$. $\quad\square$

Now back to the main proof. By (2) of the lemma, $\exists F_n \subset E_n$ which is closed, and $m^*(E_n \setminus F_n) < \frac{\varepsilon}{2^n}$. But E_n is bounded by assumption, so F_n is compact. E_n is pairwise disjoint, so F_n also is. Then

$$m^*(E_n) \leq m^*(F_n) + m^*(E_n \setminus F_n)$$

So

$$\sum_{n \geq 1} m^*(E_n) \leq \sum_n m^*(F_n) + \varepsilon \sum_{n \geq 1} \frac{1}{2^n} = m^*(\cup_n F_n) + \varepsilon \leq m^*(\cup E_n) + \varepsilon$$

But ε is arbitrary. So

$$\sum_{n\geq 1} m^*(E_n) \leq m^*(\cup_1^\infty E_n)$$

and equality follows, because the other inequality holds by countable subadditivity.

For the general case we reduce to the bounded case. We're running short of time, but the idea is that $\mathbb{R}^d = \cup_{m\geq 1}A_m$ for A_m compact. We then apply the previous case to the countable family $(A_m \cap E_n)_{n,m}$ and check that it works.

Proposition. On $(\mathbb{R}^d, \mathcal{L})$, there is a unique measure μ which is translation invariant.

The countable additivity here is crucial. There are lots of finitely additive measures which are translation invariant and unrelated to m, even more, there are such μ with $\mu(\mathbb{R}^d) < \infty$.

We'll skip the proof of this now (though it's possible to prove this directly) – we will see another proof later. $\square$

Last week we show that $\mathcal{L}$ is a σ-algebra, and the restriction of the Lebesgue outer measure, m, to $\mathcal{L}$ is a measure.

We've also shown that if $E \subseteq \mathbb{R}^d$ is Lebesgue measurable, then $\forall \varepsilon > 0 \exists U$ open set, $E \subseteq U$ s.t. $m(U \setminus E) < \varepsilon$. Also, $\forall \varepsilon > 0, \exists F$ closed set, $F \subset E$ s.t. $m(E \setminus F) < \varepsilon$.

A consequence is that, for every Lebesgue measurable set, there exists a sequence U_n of open sets, $U_{n+1} \subseteq U_n$, $E \subseteq U_n$ s.t. $m(\bigcap U_n \setminus E) = 0$. So $E = \bigcap_n u_n \setminus N$ where N is a null set ($m(N) = 0$).

Remark. A countable intersection of open sets is called a G_δ-set.

Dually, if $E \in \mathcal{L}$, $\exists F_n$ closed, $F_n \subseteq E$, s.t. $m(E \setminus \bigcup_n F_n) = 0$. So $E = \bigcup_n F_n \cup N$ where N is a null set. As a result, every Lebesgue measurable set is the countable union of a F_σ-set and a null set, where F_σ means a countable union of closed sets.

Recall that a Boolean algebra is a family of subsets of a set X, which is closed under (finite) intersection and complements and contains ϕ and X, and a σ-algebra is a boolean algebra which is also closed under countable union.

Remark. Every finite boolean algebra is a σ-algebra.

Example. We have the trivial boolean algebra, $B = \{\phi, X\}$, the discrete boolean algebra $B = \mathcal{P}(X)$, which are both also σ-algebras. We also have the elementary boolean algebra, $B = \{$finite unions of boxes in $\mathbb{R}^d$, or complements of such$\}$.

Here is a less trivial example: let X be a topological space, $B = \{$finite unions of sets of the form $U \cap F\}$, where U is open and F is closed. This is called the boolean algebra of constructible subsets of X.

Also we have the null algebra: $B = \{E \subseteq \mathbb{R}^d, m^*(E) = 0 \text{ or } m^*(E^c) = 0\}$.

Remark. The intersection of any number of boolean algebras on the same set X is again a boolean algebra, and the same holds for $\sigma-$algebras (check).

Definition. If $\mathcal{F}$ is a family of subsets of X, the boolean algebra *generated* by $\mathcal{F}$ is the intersection of all boolean algebras containing $\mathcal{F}$ (something like the minimum that contains it?). The same applies for σ-algebras: notation $\sigma(\mathcal{F})$ for the σ-algebra generated by $\mathcal{F}$.

Remark. The boolean algebra generated by a family $\mathcal{F}$ of subsets of X is the family of sets of the form

$$\bigcup_{finite} \bigcap_{finite} F \text{ or } F^c$$

for $F \in \mathcal{F}$.

Note: a finite intersection of a finite union of sets from a family $\mathcal{F}$ is always a finite union of finite intersections of sets from $\mathcal{F}$. **But this is no longer true if you change finite by countable**. For example,

$$\bigcap_{i \in \mathbb{N}} \bigcup_{j \in \mathbb{N}} E_j^{(i)} = \bigcup_{i \to j_i, \mathbb{N} \to \mathbb{N}} \bigcap_{i \in \mathbb{N}} E_{ji}^{(i)}$$

Note that there are uncountably many functions $\mathbb{N} \to \mathbb{N}$ (check Borel hierarchy), so LHS can't be equal to just interchanging the order of intersection and union.

Definition. The σ-algebra generated by the family of boxes in $\mathbb{R}^d$ is called the *Borel σ-algebra*.

Remark. It is also the σ-algebra generated by all open subsets of $\mathbb{R}^d$ (also, by all closed subsets). This is because we've shown that every open set is a countable union of boxes).

Elements of the Borel σ-algebra are called Borel set.

Remark. More generally, if X is an arbitrary topological space, the Borel σ-algebra of X is the σ-algebra generated by open subsets of X. We denote it by $\mathcal{B}(X)$.

Note that we've shown that every open subset $E \subseteq \mathbb{R}^d$ is Lebesgue measurable. Hence, $\mathcal{B}(\mathbb{R}^d) \subseteq \mathcal{L}$. The question is, are they equal?

The answer is no. One can show that $\text{Card}\,\mathcal{B}(\mathbb{R}^d) = 2^{\mathbb{N}}$, but remember that every null set lies in $\mathcal{L}$, and the middle-thirds Cantor set C is a null set $\subseteq \mathcal{L}$. We know $\text{Card}\,C = \text{Card}\{0,1\}^{\mathbb{N}} = 2^{\text{Card}\,\mathbb{N}}$, so $\text{Card}\,\mathcal{P}(C) = 2^{2^{\text{Card}\,C}}$. Also all subsets of C are null, hence in $\mathcal{L}$. So $\text{Card}\,\mathcal{L} > \text{Card}\,\mathcal{B}$.

Now let's talk about measures. Recall a measure $m(X, \mathcal{A})$, where X is a set and $\mathcal{A}$ is a σ-algebra, is a map $\mu : \mathcal{A} \to [0, +\infty]$ which is σ-additive and that s.t. $\mu(\phi) = 0$.

Null sets of $(X, \mathcal{A}, \mu)$ are subsets in $\mathcal{A}$ with μ-measure 0.

The sub-null sets are the subsets of X contained in a null set. Note that these sets might not be in $\mathcal{A}$.

Definition. The family $\mathcal{A}^*$ of subsets of X of the form $A \cup N$ or $A \setminus N$, where $A \in \mathcal{A}$, and N is a sub-null set (w.r.t. to the measure μ) forms a σ-algebra, called the *completion* of $\mathcal{A}$ w.r.t. μ. If all subnull sets are in $\mathcal{A}$, $\mathcal{A}$ is called *complete*.

Example. The completion of the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ is $\mathcal{L}$.

Basic application of measures on arbitrary measure space:

Proposition. Let $(X, \mathcal{A}, \mu)$ is a measure space. Then we have:
• upwards monotone convergence for sets: if $E_1 \subseteq E_2 \subseteq ... \subseteq E_n \subseteq ...$ for $E_n \in \mathcal{A}$, then $\lim_n \mu(E_n) = \sup \mu(E_n) = \mu(\bigcup_n E_n)$.
• downwards monotone convergence: if $E_1 \supseteq E_2 \supseteq ... \supset E_n \supseteq ...$ for $E_n \in \mathcal{A}$, if $\mu(E_1) < \infty$ then $\lim_{n \to \infty} \mu(E_n) = \inf_n \mu(E_n) = \mu(\bigcap_n E_n)$.

Proof. This follows from countable additivity of μ, say for (b) we consider the disjoint union

$$E_1 = \bigcup_{i \geq 1} E_i \setminus E_{i+1} \cup \bigcap_n E_n$$

(I think the $\cup$ and $\cap$ here is messed up – need to check) (Actually it looks correct) $\qquad \square$

Today we'll discuss how to extend a measure on Boolean algebra to the σ-algebra it generates (recall σ-algebra requires countable additivity).

To be able to extend a measure μ on a Boolean algebra $\mathcal{B}$ to $\sigma(\mathcal{B}, \mu$ has to satisfy the following property:
• It has to be σ-additive, meaning if $B_n \in \mathcal{B}$ disjoint and $\cup B_n \in \mathcal{B}$ then $\mu(\cup B_n) = \sum_{n \geq 1} \mu(B_n)$.

The theorem we are going to prove is that this is enough.

Definition. A measure μ on X is called *σ-finite* if X is a countable union of subsets with finite μ-measure.

Theorem. (Caratheodory extension theorem)
If $\mathcal{B}$ is a boolean algebra of subsets of X and μ is a countably additive measure on $\mathcal{B}$, then μ extends to a measure on $\sigma(\mathcal{B})$, which is the σ-algebra generated by $\mathcal{B}$. Moreover, the extension is unique if μ is assumed σ-finite.

Proof. We first prove existence, then prove uniqueness.

• Existence: We define for an arbitrary subset of $A \subseteq X$ the following quantity

$$\mu^*(A) = \inf \left\{ \sum_{n \geq 1} \mu(B_n), A \subseteq \bigcup_{n \geq 1} B_n \text{ and each } B_n \in B \right\}$$

Note (check!!) that $\mu^*(\phi)$ is 0, $\mu^*(A) \leq \mu^*(B)$ if $A \subseteq B$, and $\mu^*(\cup_{n \geq 1} A_n) \leq \sum_{n \geq 1} \mu^*(A_n)$ – i.e. countable subadditivity. Note that μ^* must exist because there is at least one cover, namely X itself (A set function with these 3 properties is called an *outer-measure*). We introduce the following auxiliary definition:

Definition. A subset $E \subseteq X$ will be called μ^*-measurable (or Caratheodory measurable w.r.t. μ^*) if for every subset $A \subseteq X$ (does it need to be in $\mathcal{B}$?), $\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$.

We claim that (Claim 1) the family $\mathcal{B}^*$ of μ^*-measurable subsets is a σ-algebra, and μ^* is a measure on $\mathcal{B}^*$. Also, we claim that (Claim 2) $\mu^*(B) = \mu(B)$ for all $B \in \mathcal{B}$, and $\mathcal{B} \subseteq \mathcal{B}^*$ – therefore we have proven existence.

• Claim 1: we first subclaim that $\mathcal{B}^*$ is a Boolean algebra. It's clear that $\mathcal{B}^*$ is closed under complementation (by definition): $E \in \mathcal{B}^* \implies E^c \in \mathcal{B}^*$. Also $\mathcal{B}^*$ contains ϕ and X; it remains to show that $\mathcal{B}^*$ is closed under finite unions.

Let $E, F \in \mathcal{B}^*$, and $A \subseteq X$ be an arbitrary subset. Now

$$E \in \mathcal{B}^* \implies \mu^*(A \cap (E \cup F)) = \mu^*(A \cap (E \cup F) \cap E) + \mu^*(A \cap (E \cup F) \cap E^c)$$
$$= \mu^*(A \cap E) + \mu^*(A \cap (F \setminus E))$$
$$F \in \mathcal{B}^* \implies \mu^*(A \cap E^c) = \mu^*(A \cap E^c \cap F) + \mu^*(A \cap E^c \cap F^c)$$
$$= \mu^*(A \cap (F \setminus E)) + \mu^*(A \cap (E \cup F)^c)$$

add the two previous equalities we get

$$\mu^*(A \cap (E \cup F)) + \mu^*(A \cap (E \cup F)^c) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$$
$$= \mu^*(A)$$

So $E \cup F \in \mathcal{B}^*$, i.e. $\mathcal{B}^*$ is a boolean algebra.

Now let's prove that $\mathcal{B}^*$ is a σ-algebra. Note (check!) that since $\mathcal{B}^*$ is a boolean algebra, it is enough to check that $\cup_n E_n \in \mathcal{B}^*$ for any coutnable pairwise disjoint family $(E_n)_n$ of subsets of $\mathcal{B}^*$. Let $A \subseteq X$ be an arbitrary subset, then

$$E_1 \in \mathcal{B}^* \implies \mu^*(A) = \mu^*(A \cap E_1) + \mu^*(A \cap E_1^c)$$
$$E_2 \in \mathcal{B}^* \implies \mu^*(A \cap E_1^c) = \mu^*(A \cap E_1^c \cap E_2) + \mu^*(A \cap E_1^c \cap E_2^c)$$

Keep going this way, we get

$$\mu^*(A \cap E_1^c \cap E_2^c \cap ... \cap E_n^c) = \mu^*(A \cap E_{n+1}) + \mu^*(A \cap E_1^c \cap ... \cap E_{n+1}^c)$$

Then add these up, and we get for every integer n,

$$\mu^*(A) = \sum_{n=1}^{N} \mu^*(A \cap E_n) + \mu^*(A \cap E_1^c \cap ... \cap E_N^c)$$

In particular, $\mu^*(A) \geq \sum_{n=1}^{N} \mu^*(A \cap E_n) + \mu^*(A \cap (\bigcup_n E_n)^c)$. Now we can let $N \to \infty$ we get

$$\mu^*(A) \geq \sum_{1}^{\infty} \mu^*(A \cap E_n) + \mu^*(A \cap (\cup E_n)^c)$$

However,

$$\mu^*(A) \leq \mu^*(A \cap \bigcup E_n) + \mu^*(A \cap (\bigcup E_n)^c)$$

$$\leq \sum_{n=1}^{\infty} \mu^*(A \cap E_n) + \mu^*(A \cap (\bigcup E_n)^c)$$

by countable subadditivity of μ^*. Hence in fact $\mu^*(A) = \mu^*(A \cap \bigcup E_n) + \mu^*(A \cap (\bigcup E_n)^c)$, i.e. $\bigcup_n E_n \in \mathcal{B}^*$; on the other hand, from $\mu^*(A \cap \bigcup E_n) = \sum_{n=1}^{\infty} \mu^*(A \cap E_n)$ we get μ^* is countably additive on $\mathcal{B}^*$. This ends the proof for claim 1, i.e. Tthe family $\mathcal{B}^*$ of μ^*-measurable subsets is a σ-algebra and μ^* is a measure on $\mathcal{B}^*$.

• Claim 2: We need to show $B \in \mathcal{B}^*$ for all $B \in \mathcal{B}$. Pick an arbitrary $A \subseteq X$. We need to show that $\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c)$. It's clear by subadditivity of μ^* that LHS $\leq$ RHS. To prove LHS $\geq$ RHS, by definition of μ^*, $\forall \varepsilon > 0$ $\exists B_n \in \mathcal{B}, A \subseteq \bigcup_{n=1}^{\infty} B_n$ s.t. $\mu^*(A) + \varepsilon \geq \sum_{n=1}^{\infty} \mu^*(B_n)$. We write

$$A \cap B \subseteq \bigcup_n B_n \cap B \in \mathcal{B}$$

and

$$A \cap B^c \subseteq \bigcup_n B_n \cap B^c \in \mathcal{B}$$

so we get

$$\mu^*(A \cap B) \leq \sum \mu(B_n \cap B)$$
$$\mu^*(A \cap B^c) \leq \sum \mu(B_n \cap B^c)$$

But μ is additive on $\mathcal{B}$, so $\mu(B_n) = \mu(B_n \cap B) + \mu(B_n \cap B^c)$. So

$$\mu^*(A \cap B) + \mu^*(A \cap B^c) \leq \sum \mu(B_n \cap B) + \mu(B_n \cap B^c) = \sum \mu(B_n)$$
$$\leq \varepsilon + \mu^*(A)$$

i.e. $B \in \mathcal{B}^*$. So $\mathcal{B} \subseteq \mathcal{B}^*$.

It remains to show that $\mu^*(B) = \mu(B)$ if $B \in \mathcal{B}$. Again LHS $\leq$ RHS is clear (check) by definition of μ^*, and conversely, $\forall \varepsilon > 0$, $\exists B_n \in B$, $B \subseteq \bigcup B_n$ s.t. $\varepsilon + \mu^*(B) \geq \sum_n \mu(B_n)$. But

$$\mu(B) \leq \mu(\bigcup B_n) \leq \sum \mu(B_n)$$
$$\leq \mu^*(B) + \varepsilon$$

So μ^* is a measure on $\sigma(\mathcal{B}) = \mathcal{B}^*$, i.e. we've proven existence.

• Proof of uniqueness: μ is σ-finite, i.e. $X = \bigcup_n E_n$ with $\mu(E) < \infty$. $\mu|_{E_n}$ is a finite measure on E_n, and so WLOG we may assume that $\mu(x) < \infty$.

Let ν be a measure on $\sigma(\mathcal{B})$ extending μ, i.e. $\nu(B) = \mu(B) \ \forall B \in \mathcal{B}$. We need to show that $\nu(E) = \mu^*(E) \forall E \in \sigma(\mathcal{B})$. For this, consider ν^* defined as before with $(v, \sigma(\mathcal{B}))$ in place of $(\mu, \mathcal{B})$. We've proved that $\nu^* = \nu$ on $\sigma(\mathcal{B})$. On the other hand, $\nu^*(A) \leq \mu^*(A) \ \forall A \subseteq X$ by definition of ν^* and μ^*. So if $A \in \sigma(\mathcal{B})$

then $\nu(A) \leq \mu^*(A)$. Hence $\nu(A^c) \leq \mu^*(A^c)$. (????) But $\nu(A^c) = \nu(X) - \nu(A)$, $\mu^*(A^c) = \mu^*(X) - \mu^*(A)$, so $\nu(A) \geq \mu^*(A)$.

$\square$

Exercise: Recall that $\mathcal{B}^* = \{E \subseteq X, E$ is μ^*-measurable $\}$ (recall that μ^*-measurable means $\mu^*(A) + \mu^*(A \cap E) + \mu^*(A \cap E^c))\forall A \subseteq X)$.
• Every subset in $\mathcal{B}^*$ is of the form $E = A \cup N$ where $A = \sigma(B)$ and $\mu^*(N) = 0$. So in fact, $\mathcal{B}^*$ is the completion of $\sigma(\mathcal{B})$ wrt μ^*.
• When $\mathcal{B}$ = the elementary boolean algebra of $\mathbb{R}^d$ (= finite unions of boxes or complements of such), then $\sigma(\mathcal{B})$ = Borel σ-algebra,$\mathcal{B}^* = \mathcal{L} = \sigma$-algebra of Lebesgue measurable sets.

Remark. This means that the Caratheodory extension theorem gives an alternative way to construct the σ-algebra of Lebesgue measurable sets in $\mathbb{R}^d$, and the Lebesgue measure.

There is a much stronger uniqueness theorem.

Theorem. Suppose $(X, \mathcal{A})$ is a measurable space, and μ_1 and μ_2 are two finite measures on $(X, \mathcal{A})$ such that $\mu_1(X) = \mu_2(X)$. Suppose that $\mu_1(A) = \mu_2(A)$ $\forall A \in \mathcal{F}$, where $\mathcal{F}$ is a subfamily of $\mathcal{A}$ which generates $\mathcal{A}$ (i.e. $\sigma(\mathcal{F}) = \mathcal{A}$), and is stable under finite intersections. Then $\mu_1 = \mu_2$.

Lemma. (Dynkin's)
Let $\mathcal{A}$ be a σ-algebra and $\mathcal{F}$ a family of subsets on $\mathcal{A}$, which generates $\mathcal{A}$ and is stable under finite intersections and contains ϕ (such $\mathcal{F}$ is sometimes called a π-system). Let $\mathcal{C}$ be a subfamily of $\mathcal{A}$, which contains $\mathcal{F}$, and is stable under complementation and countable disjoint unions. Then $\mathcal{C} = \mathcal{A}$.

Proof of uniqueness theorem via Dynkin's lemma: Observe that $\mathcal{C} = \{A \in \mathcal{A}, \mu_1(A) = \mu_2(A)\}$. Then $\mathcal{C}$ satisfies the assumptions of Dynkin's lemma, and so $\mathcal{C} = \mathcal{A}$. So $\mu_1 = \mu_2$.

1.3 Measurable Functions

Definition. $(X, \mathcal{A})$ measurable space, and $f : X \to \mathbb{R}$ is a real valued function. Say that f is $\mathcal{A}$-measurable (or just measurable if it's clear which σ-algebra we are talking about) if $\forall t \in \mathbb{R}$, $\{x \in X, f(x) < t\} \in \mathcal{A}$.

Remark. If f is $\mathcal{A}$-measurable, then $f^{-1}(B) \in \mathcal{A}$ for all Borel subsets $B \subseteq \mathbb{R}$. Indeed f^{-1} is 'functional' w.r.t. $\cup, \cap$ and complement. So $\{B \subseteq \mathbb{R}|f^{-1}(B) \in \mathcal{A}$ is a σ-algebra. If f is measurable, by definition it contains $(-\infty, t)$ $\forall t$. So it contains all intervals, hence all Borel sets.

Caveat: The pre-image under f of a Lebesgue measurable subset of $\mathbb{R}$ (which is not Borel) may not be in $\mathcal{A}$. More generally, if $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ are measurable spaces, then

Definition. $f : X \to Y$ is said to be measurable if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$.

Remark. To check that $f : X \to Y$ is measurable, it is enough to check that $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{F} =$ generating subfamily of $\mathcal{B}$.

Remark. Let $f : X \to Y$, $g : Y \to Z$, and we have $(X, \mathcal{A}), (Y, \mathcal{B}), (Z, \mathcal{C})$. If f, g are measurable, so is $f \circ g : X \to Z$ (obvious).

Remark. $(X, \mathcal{A})$ a measurable space, $f : X \to \mathbb{R}^d$ is $\mathcal{A}$-measurable if and only if each coordinate function f_i, $i = 1, ..., d$ is $\mathcal{A}$-measurable, where $f = (f_1, ..., f_d) \in \mathbb{R}^d$.

Proposition. If f, g are $\mathcal{A}$-measurable $X \to \mathbb{R}$ function, then so is f_g, fg, df for $d \in \mathbb{R}$. In particular, the set of $\mathcal{A}$-measurable functions on X is a vector space stable under multiplication ($\mathbb{R}$-algebra).

Proof. Consider $X \to \mathbb{R}^2 \xrightarrow{s} \mathbb{R}$ by $x \to (f(x), g(x)) \to f(x) + g(x)$, where $s : \mathbb{R}^2 \to \mathbb{R}$ by $(x, y) \to x + y$. By previous remark, F is measurable. But s is measurable because $\{(x, y) \in \mathbb{R}^2, x + y < t\}$ is open and therefore Borel. Hence the composition $s \circ f$ is measurable. $\square$

Remark. Every continuous function $\mathbb{R} \to \mathbb{R}$ is measurable. Why? because $\{x, f(x) < t\}$ is open, hence Borel.

Proposition. If $(f_n)_n$ is a family of measurable functions $X \to \mathbb{R}$, then $\limsup_{n \to \infty} f_n$, $\liminf_{n \to \infty} f_n$, $\sup f_n$, $\in f_n$ are all $\mathcal{A}$-measurable.

Proof. Consider $\{x \in X; \liminf_n f_n < t\} = \bigcup_{k \geq 1} \bigcap_{m \geq 1} \bigcup_{n \geq m} \{x \in X, f_n(x) < t - 1/k\}$. The others are similar (wtf). n $\square$

Remark. If $X = \bigcup_{i=1}^N P_i$ is a partition of X, and $\mathcal{A}$ is the atomic boolean algebra associated with this partition (i.e. subsets of X belonging to $\mathcal{A}$ are the finite unions of P_i), then $f : X \to \mathbb{R}$ is $\mathcal{A}$-measurable iff f is constant on each P_i for $i = 1, ..., N$ iff $f = \sum_{i=1}^N a_i 1_{P_i}$ for $a_i \in \mathbb{R}$.

The intuition is that f is $\mathcal{A}$-measurable means that $f(x)$ depends only on which subsets from $\mathcal{A}$ x belongs to.

Remark: $E \in \mathcal{A}$ iff 1_E is $\mathcal{A}$-measurable, for every subset $E \subseteq X$.

Definition. If X is a topological space, a Borel function on X is a function which is
$mathcalB(X)$-measurable, where $\mathcal{B}(X)$ is the Borel σ-algebra, i.e. the σ-algebra generated by the open sets.

Caveat: if $f : X \to \mathbb{R}$ is $\mathcal{A}$-measurable, then $f^{-1}(B) \in \mathcal{A}$ for all Borel measurable subsets $B \subseteq \mathbb{R}$. But $f^{-1}(E)$ for E Lebesgue measurable may not be in A.

Remark. (The devil's staircase)
Consider a non-decreasing function $\phi : [0, 1] \to [0, 1]$ that is continuous, and $\phi(0) = 0$, $\phi(1) = 1$, and is constant on each component of $[0, 1] \setminus C$, where C is the middle-thirds Cantor set. We can show that ϕ converges uniformly to its pointwise limit.

Exercise: show that there exists a continuous increasing bijection $\phi : [0, 1] \to [0, 1]$, $\phi(0) = 0$, $\phi(1) = 1$, and a subset $F \subseteq [0, 1]$ of Lebesgue measure zero, such that $phi(F)$ has positive measure. (consider allowing a small positive slope on each component of the $[0, 1] \setminus C$ in the above remark)

Exercise: show that the above exercise gives rise to a counter example explaining the above Caveat.
• hint: show that there exists $A \subseteq \phi(F)$ non Lebesue measurable. Set $f = \phi^{-1}$, $E = f(A)$.
• hint: show that every subset of $\mathbb{R}$ of positive measure contains a non Lebesgue measurable subset (modify Vitali's construction in the first chapter).

1.4   Integration

Let $(X, \mathcal{A}, \mu)$ be a measure space. We are going to define the integral $\mu(f) = \int_X f d\mu = \int_X f(x) d\mu(x)$ of a function f w.r.t. μ.

First we define $\mu(f)$ when f is a *simple function*, i.e. a function of the form $f = \sum_{i=1}^{N} a_i 1_{E_i}$, a linear combination of indicator functions with each $E_i \in \mathcal{A}$. We define $\mu(f) := \sum_{i=1}^{N} a_i \mu(E_i)$.

Check as an exercise, that this definition makes sense (it's independent of the representation of f).

Now if $f : X \to \mathbb{R}$ has values in $[0, +\infty]$ and is $\mathcal{A}$-measurable, we define

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

where by $g \leq f$ we mean that $g(x) \leq f(x)$ for all x.

Finally, if $f : X \to \mathbb{R}$ is any $\mathcal{A}$-measurable function, then we set $f^+(x) := \max\{0, f(x)\}$ and $f^- := (-f)^+$. Then clearly $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

Definition. f is μ-integrable if $\mu(f^+) < \infty$ and $\mu(f^-) < \infty$, and if this is the case we define
$$\mu(f) = \mu(f^+) - \mu(f^-)$$

Theorem. (Monotone convergence theorem)
Let $(f_n)_n$ be a sequence of $\mathcal{A}$-measurable functions, such that for all $x \in X$, $0 \leq f_1(x) \leq f_2(x) \leq \dots \leq f_n(x) \leq \dots$. Let

$$f(x) = \lim_{n \to \infty} f_n(x) \in [0, +\infty]$$

Then

$$\mu(f_n) \xrightarrow{n \to \infty} \mu(f)$$

Proof. Clearly $\mu(f_n) \leq \mu(f)$ for all n. Also $\mu(f_n) \leq \mu(f_{n+1})$ for all n. For all $\varepsilon > 0$, let g be a simple function, $g \leq f$, and let $E_n = \{x \in X; f_n(x) \geq (1 - \varepsilon)g(x)\}$. Note $E_n \in \mathcal{A}$, and $E_n \subseteq E_{n+1}$ and $X = \bigcup_{n \geq 1} E_n$. So we can apply the upwards monotone convergence for sets to the measure $m_g : E(\in \mathcal{A}) \to \mu(1_E \cdot g)$ (check

that this is a measure on $(X, \mathcal{A}$. It's a product inside the bracket), and get $m_g(E_n) \uparrow m_g(X) = \mu(g)$ as $n \to +\infty$ (what's up arrow – converge monotonically from below?). Now

$$m_g(E_n) = \int 1_{E_n}(x) \cdot g(x) d\mu(x)$$
$$\leq \frac{1}{1-\varepsilon} \int f_n(x) d\mu(x)$$
$$= \frac{\mu(f_n)}{1-\varepsilon}$$

So we get $\mu(g) \leq \frac{1}{1-\varepsilon} \liminf_n \mu(f_n) \leq \limsup_n \mu(f_n) \leq \mu(f)$.

Now this holds for all g simple and $0 \leq g \leq f$, so taking the sup over all such g, we get

$$\mu(f)(1-\varepsilon) \leq \liminf \leq \limsup \leq \mu(f)$$

(fill what's after liminf and limsup)
This holds for all $\varepsilon > 0$, so we've got

$$\mu(f) = \liminf \mu(f_n) = \limsup \mu(f_n).$$

$\square$

Corollary. f, g are μ-integrable functions on $(X, \mathcal{A}, \mu)$, then
• for $a, b \in \mathbb{R}$, $\alpha f + \beta g$ is μ-integrable, and $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$;
• if $f \geq 0$, then $\mu(f) \geq 0$;
• if $f \geq 0$ and $\mu(f) = 0$, then $f = 0$ *a.e.*.

Here *a.e.* means 'almost everywhere': a property $P(x)$ holds a.e. means that the set on which it doesn't hold is a (sub)-null set for μ.

Proof. • We know the linearity for simple functions, so we will reduce to this case using the Monotone Convergence Theorem.

First, writing $f = f^+ - f^-$, $g = g^+ - g^-$ where f^+, f^- are as defined previously, we reduce to the case when both f and g are non-negative; finally every non-negative measurable function is a pointwise limit of simple functions: indeed let $f_n(x) := \min(\frac{1}{n^2}[n^2 f(x)]))$, where the square bracket is the floor function. This function only takes finitely many values, and also $f_n(x) \leq f(x)$ $\forall n$ and x, and $\lim_{n\to\infty} f_n(x) = f(x)$. So now just apply the monotone convergence theorem to the sequence (f_n).

• Recall $\mu(f)$ was defined as $\sup\{\mu(g), 0 \leq g \leq f, g \text{ simple}\}$. So it's clear that $f \geq 0 \implies \mu(f) \geq 0$.

• Now if $f \geq 0$ and $\mu(f) = 0$, consider $E_n = \{x \in X, f_n(x) \geq \frac{1}{n}\}$. Note $E_n \in \mathcal{A}$, and $\bigcup_{n\geq 1} E_n \cup \{x \in X, f(x) = 0\} = X$. By countable subadditivity of μ, it is enough to show that $\forall n$, $\mu(E_n) = 0$. But $\frac{1}{n} \cdot 1_{E_n} \leq f$, so $\frac{1}{n}\mu(E_n) \leq \mu(f)$. As $\mu(f) = 0$, we get $\mu(E_n) = 0$. $\square$

Remark. If f and g are measurable and integrable, and $f = g$ a.e., then g is integrable and $\mu(f) = \mu(g)$. (check)

Theorem. (Fatou's lemma)

If $(f_n)_n$ is a sequence of measurable functions, $m(X, \mathcal{A}, \mu)$ such that $f_n \geq 0 \; \forall n$, then

$$\mu(\liminf_{n\to\infty} f_n) \leq \liminf_{n\to\infty} \mu(f_n)$$

Note that this inequality can be strict. For example, in the following three types of 'moving bumps' examples:

(a) $X = [0, +\infty)$ with Lebesgue measure, $f_n = 1_{[n,n+1]}$, then $\lim f_n = 0$ but $\mu(f_n) = 1$ for all n.

(b) Take $f_n = \frac{1}{n} 1_{[0,n]}$, then $\liminf f_n = 0$ but $\mu(f_n) = 1$ for all n again.

(c) Let $X = [0, 1]$ with Lebesgue measure, and take $f_n = 1_{[\frac{1}{n}, \frac{2}{n}]} \cdot n$. Again, $\lim f_n = 0$ but $\mu(f_n) = 1$.

Proof. (See an alternative proof that doesn't rely on monotone convergence theorem here).

Let $g_n = \inf_{k \geq n} f_k$. Note $g_{n+1} \geq g_n \geq 0$, so $\mu(g_n) \xrightarrow{n\to\infty} \mu(g_\infty)$ where $g_\infty = \lim_{n\to\infty} g_n$; $g_n \leq f_n$, so $\mu(g_n) \leq \mu(f_n)$ i.e. $\lim \mu(g_n) \leq \liminf_n \mu(f_n)$. Therefore

$$\mu(g_\infty) \leq \liminf_n \mu(f_n)$$

But $g_\infty = \liminf_{n\to\infty} f_n$. $\qquad\square$

Theorem. (Lebesgue's dominated convergence theorem)

Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $(f_n)_n$ be a sequence of measurable functions on X. Assume that there exists a μ-integrable functions g such that $|f_n| \leq g$ for all n ('dominated' by g). Also, assume further that $f_n(x)$ converges pointwise to $f(x)$ for all $x \in X$. Then we can interchange the order of integration and limit, i.e.

$$\mu(\lim_{n\to\infty} f_n) = \lim_{n\to\infty} \mu(f_n)$$

Proof. Note $|f_n| \leq g \implies |f| \leq g$, where $f = \lim_{n\to\infty} f_n$. So f is integrable, and $\mu(|f|) \leq \mu(g) < \infty$. Now observe that $g + f_n$ is non-negative. So apply Fatou's lemma to $(g + f_n)_n$, we get

$$\mu(\liminf_{n\to\infty} g + f_n) \leq \liminf \mu(g + f_n)$$

But also LHS $= \mu(g + f) \leq \mu(g) + \liminf_n \mu(f_n)$. Cancelling $\mu(g)$ we get $\mu(f) \leq \liminf \mu(f_n)$. Now do the same with $-f_n$ in place of f_n, we get $\mu(-f) \leq \liminf \mu(-f_n)$, i.e. $\mu(f) \geq \limsup \mu(f_n)$. So

$$\mu(f) \leq \liminf \mu(f_n) \leq \limsup \mu(f_n) \leq \mu(f)$$

Hence $\mu(f_n)$ converges to $\mu(f)$. $\qquad\square$

Corollary. (exchange of $\sum$ and $\int$)

• Let $(f_n)_{n\geq 1}$ be a sequence of measurable sets. Assume $f_n \geq 0$ for all n. Then $\mu(\sum_{n\geq 1} f_n) = \sum_{n\geq 1} \mu(f_n)$.

• Let $(f_n)_{n\geq 1}$ for any sequence of measurable functions. Assume $\sum_{n\geq 1} |f_n|$ is $\mu-$integrable. Then we can exchange $\sum$ and $\int$, i.e.

$$\mu(\sum_{n\geq 1} f_n) = \sum_{n\geq 1} \mu(f_n)$$

Proof. • Just take $g_N = \sum_{n=1}^{N} f_n$, the partial sum, and apply the Monotone Convergence Theorem to $(g_N)_{N \geq 1}$.

• Let $g = \sum_{n \geq 1} |f_n|$ and again $g_N = \sum_{n=1}^{N} f_n$. Then $\forall N$, $|g_N| \leq g$, g is μ-integrable by assumption. Then just apply the Lebesgue Dominated Convergence Theorem. $\qquad\square$

Corollary. (exchange of $\frac{\partial}{\partial x}$ and $\int$)
Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $f : I \times X \to \mathbb{R} : (t, x) \to f(t, x)$ where I is an open interval in $\mathbb{R}$. Assume
• $x \to f(t, x)$ is μ-integrable $\forall t \in I$;
• $t \to f(t, x)$ is differentiable $\forall x \in X$;
• (domination) there exists a function $g : X \to \mathbb{R}$ s.t. $|f(t, x) \leq g(x)|$ for every $x \in X, t \in I$, and g is μ-integrable.

Then $x \to \frac{\partial f}{\partial x}(t, x)$ is μ-integrable $\forall t \in I$, $t \to \int_X f(t, x) d\mu$ is integrable, and

$$\frac{\partial}{\partial t} \left(\int_X f(t, x) d\mu(x) \right) = \int_X \frac{\partial f}{\partial t}(t, x) d\mu(x)$$

Proof. This is a corollary of the Lebesgue Dominance Convergence Theorem.

Let $h_n \in \mathbb{R}$, $h_n \xrightarrow{n \to \infty} 0$. Set $g_n(x) := \frac{f(t+h_n, x) - f(x)}{h_n}$. Then $g_n(x) \to \frac{\partial f}{\partial t}$ pointwise. But there exists $\theta_n \in [t, t + h_n]$ s.t. $g_n(x) = \frac{\partial f}{\partial t}(t, x)|_{t=\theta_n}$ by mean value theorem. So

$$|g_n| \leq g$$

and g is integrable. So apply LDC Theorem. $\qquad\square$

In the proof of the monotonic convergence theorem, we've used the fact that if g is a simple function, then $A \in \mathcal{A} \to \mu(g 1_A)$ is a measure. Now this is *clear* reducing to the case $g = 1_B$ by linearity. And it also holds for all $g \geq 0$ measurable, but this is an exercise which may require the monotone convergence theorem.

Theorem. (Egoroff Theorem)
Let $(X, \mathcal{A}, \mu)$ be a measure space. Assume $\mu(X) < \infty$, $(f_n)_{n \geq 1}$ a sequence of measurable facts which converges pointwise to $f = \lim f_n$. Then $\forall \varepsilon > 0$ $\exists A \in \mathcal{A}$, $\mu(A) > \mu(X) - \varepsilon$ s.t.

$$\sup_{x \in A} |f_n(x) - f(x)| \xrightarrow{n \to \infty} 0$$

Proof. Replacing f_n by

$$\sup_{k \geq n} |f_k(x) - f(x)|$$

we may assume that $f = 0$, $0 \leq f_{n+1} \leq f_n$. Then $f_n(x) \to 0$ for all $x \in X$, so $\forall k \exists n_k(x) \in \mathbb{N}$ s.t. $0 \leq f_{n_k}(x) \leq 1/k$.

So $n_k(x) < \infty$ for all x. So $\bigcup_{1 \leq T \infty} \{x \in X, n_k(x) \leq T\} = X$ for every k. Hence $\mu\{x; n_k \leq t\} \xrightarrow{T \to \infty} \mu(x)$ by monotone convergence theorem for sets.

Let $T_k \in \mathbb{N}$ be such that $\mu\{x, n_k \leq T_k\} \geq \mu(X) - \frac{\varepsilon}{2^k}$. Take $A = \bigcap_{k \geq 1}\{x; n_k(x) \leq T_k\}$. By countable subadditivity of μ we get $\mu(X \setminus A) \leq \sum_{k \geq 1} \varepsilon/2^k \leq \varepsilon$. But we have $\sup_{x \in A} f_n(x) \xrightarrow{n \to \infty} 0$ because $\sup_A f_{T_k} \leq 1/k$ for all k, and $n \to \sup_A f_n$ is non-increasing. $\square$

Theorem. (Fundamental theorem of calculus)
$f : [a, b] \to \mathbb{R}$ is continuous, and let $F(x) = \int_{[a,x]} f d\mu$, where μ is a Lebesgue measure. Then F is differentiable, and $F'(x) = f(x)$ for all x.

Proof. Same as for the Riemann integral. This is actually easier because we assumed continuity.

Indeed $\frac{F(x+h)F(x)}{h} = \int_x^{x+h} \frac{f(t)}{h} dt$ but by continuity this converges to $f(x)$ as $h \to 0$. $\square$

Remark. There is a much more powerful version of this theorem, namely:

Theorem. Let $f : [a, b] \to \mathbb{R}$ be a μ-integrable function, and μ is a Lebesgue measure. Let $F(x) := \int_{[a,x]} f d\mu$. Then F is differentiable almost everywhere w.r.t. μ, and $F'(x) = f(x)$ μ−almost everywhere.

Theorem. (Lebesgue differentiation theorem)
Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lebesgue integrable function. Then the following holds μ-almost everywhere:

$$\frac{1}{\mu(B(x,r))} \int_{B(x,r)} f d\mu \xrightarrow{r \to 0} f(x)$$

where $B(x, r) = \{y \in \mathbb{R}^d, ||x - y|| < r\}$. The norm used here is not important.

Remark. $F'(x) = f(x)$ almost everywhere, but not everywhere. For example, take $f = 1_{[0,1]}$, then $F(x) = \min\{x, 1\}$ if $x \geq 0$; this is not differentiable at 1.

We are also interested in the converse statement as well, but it fails in general. Namely, if $F : [a, b] \to \mathbb{R}$ is continuously and differentiable almost everywhere, then it's not always true that $F(x) - F(a) = \int_a^x F'(t)dt$. For example, take $F(x) = x^2 \sin(\frac{1}{x^2})$. F' exists for all $x \in [0, 1]$, but it's not Lebesgue integrable. Another example is the Devil's staircase which we've seen before: F is a non-decreasing continuous function, but $F'(x) = 0$ for all $x \notin C$ and F is not differentiable on C, i.e. $F'(x) = 0$ almost everywhere. However, $F(0) = 0$ and $F(1) = 1$.

Remark. However, if F is Lipschitz, i.e. $\exists L > 0$ s.t. $|F(x) - F(y)| \leq L|x - y|$ $\forall x, y \in [a, b]$, then F is a.e. differentiable (this is called the Rademacher differentiation theorem). Note that tour previous converse statement will hold for such F. Indeed, $F'(x) = \lim_{n \to \infty} F_n(x)$, $F_n(x) := n(F(x + \frac{1}{n}) - F(x))$. Then by Lipschitz condition we know $|F_n(x)| \leq L$ for all x. Then we can apply the LDC theorem to swap the limit and integral:

$$\int_a^b F'(x)dx = \int_a^b \lim_n F_n(x)dx = \lim_n \int_a^b F_n(x) = F(b) - F(a).$$

Theorem. (Change of variable formula)
If $U, V \subseteq \mathbb{R}^d$ are open sets, and $\phi : U \to V$ is a C^1-diffeomorphism, then $f : V \to \mathbb{R}$ is Lebesgue measurable. Assume either $f \geq 0$ or f integrable. Then

$$\int_\mu f(\phi(x))|J_\phi(x)|dx = \int_v f(y)dy$$

where dx and dy are Lebesgue measure, and $J_\phi(x) = \det(d\phi(x))$, i.e. the Jacobian.

Remark. • We'll skip the proof (prove it for linear ϕ, then approximate on small boxes around each point).
• The theorem holds if ϕ is only assumed to be a Lipschitz homomorphism as well.
• When $u = [a, b]$, this is basically the Fundamental theorem of calculus.
• This shows that Lebesgue measure is invariant under linear affine maps $x \to Ax + b$, $b \in \mathbb{R}^d$, $A \in GL_d(\mathbb{R})$ with determinant 1.

1.5 Product measures

Definition. (Product σ-algebra)
Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. On $X \times Y$ we defined the product σ-algebra, denoted by $\mathcal{A} \otimes B$ to be the sigma algebra generated by all $A \times B \subseteq X \times Y$, with $A \in \mathcal{A}$, $B \in \mathcal{B}$.

Remark. The projection maps $\pi_X : X \times Y \to X$, $\pi_Y : X \times Y \to Y$ that maps (x, y) to x and y respectively are measurable (preimage of measurable sets are measurable), when $X \times Y$ is endowed with $A \otimes B$. $A \times \mathcal{A}$, $\pi_X^{-1} = A \times Y$.
In fact, $A \otimes B$ is the smallest $\sigma-$algebra on $X \times Y$ for which both projections are measurable.
Consider $\mathcal{B}(\mathbb{R}^{d_1+d_2}) = \mathcal{B}(\mathbb{R}^{d_1}) \otimes \mathcal{B}(\mathbb{R}^{d_2})$. In fact, if X and Y are 2nd countable topological spaces,i.e. those which have a countable base, then $\mathcal{B}(X \times Y) = \mathcal{B}(X) \otimes \mathcal{B}(Y)$, where $\mathcal{B}(X)$ is the sigma algebra generated by open sets. However, this is wrong in general.

Lemma. If $E \subseteq X \times Y$ belongs to $\mathcal{A} \otimes \mathcal{B}$, then $\forall x \in X, E_x \in \mathcal{B}$, and $\forall y \in Y, E_y \in \mathcal{A}$, where $E_x = \{y \in Y, (x, y) \in E\}$, and $E_y = \{x \in X, (x, y) \in E\}$.

Proof. Let $\mathcal{C} := \{E \subseteq X \times Y \, s.t. E_x \in \mathcal{B}, E_y \in \mathcal{A} \forall x \in X, y \in Y\}$. Check that $\mathcal{C}$ is a σ-algebra. Also, $\mathcal{C}$ contains all subsets of the form $A \times B$ for $A \in \mathcal{A}, B \in \mathcal{B}$. So we get $\mathcal{C} \supseteq \mathcal{A} \otimes \mathcal{B}$. $\square$

Using this lemma, we can prove that (as an exercise) $\mathcal{L}(\mathbb{R}^{d_1}) \otimes \mathcal{L}(\mathbb{R}^{d_2}) \subsetneq \mathcal{L}(\mathbb{R}^{d_1+d_2})$, i.e. they are not equal.

Proposition. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two $\sigma-$finite (i.e. a space can be written as a countable union of increasing subsets with finite measure) measure spaces. Then there is a unique measure on $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ denoted by $\mu \otimes \nu$ s.t. $\mu \otimes \nu(A \times B) = \mu(A) \cdot \nu(B)$ for all $A \in \mathcal{A}, B \in \mathcal{B}$. This is called the product measure.

Proof. (sketch)
Let $\mathcal{B}_0$ be the boolean algebra generated by the product sets $A \times B$, $A \in \mathcal{A}$, $B \in \mathcal{B}$. Note $\mathcal{B}_0$ is made of disjoint finite unions of product sets. The equation given in the proposition defines a measure on $\mathcal{B}_0$, and one needs to check countable additivity and then we can apply the Caratheodory extension theorem to claim the existence and uniqueness of the extension to $\sigma(\mathcal{B}_0) = \mathcal{A} \otimes \mathcal{B}$. $\qquad\square$

Theorem. (Tonnelli - Fubini theorem)
Let $(X, \mathcal{A}, \mu), (Y, \mathcal{B}, \nu)$ be σ-finite measure spaces. Let f be a $\mathcal{A} \otimes \mathcal{B}$-measurable function on $X \times Y$.
• Assume $f \geq 0$. Then the functions $f_x(y) := \int_X f(x,y)d\mu(x)$, $f^y(X) := \int_Y f(x,y)d\nu(y)$ are measurable, w.r.t. $\mathcal{B}$ and $\mathcal{A}$ respectively, and

$$\int_{X \times Y} f(x,y)d\mu \otimes \nu(x,y) = \int_Y f_x(y)d\nu(y) = \int_X f^y(x)d\mu(x) \quad (**)$$

• Now assume f is $\mu \otimes \nu$-integrable. Then f_x is ν-integrable, f^y is μ-integrable, and (**) holds.

Remark. When $f = 1_E$, $E \in \mathcal{A} \otimes \mathcal{B}$, then $x \to \nu(E_x)$, $y \to \mu(E_y)$ are measurable functions, and
$$\mu \otimes \nu(E) = \int_X \nu(E_x)d\mu = \int_Y \mu(E_y)d\nu$$

Remark. The theorem still holds if f is only assumed measurable w.r.t. the completion $(\mathcal{A} \otimes \mathcal{B})^*$, i.e. its completion.

Proof. (of theorem, sketch)
WLOG we assume that $\mu(x) < \infty$, $\nu(y) < \infty$ ($X = \bigcup E_n$, $\mu(E_n < \infty)$). Write f as $f = \lim_n f_n$ pointwise limit of simple functions (e.g. $f(x) = \lim_{n \to \infty} \frac{1}{n^2} \min([n^2 f(x)], n)$), we reduce to prove the theorem for f simple (using the monotone convergence theorem). By linearity, we can reduce tot he case when $f = 1_E$ for some $E \in \mathcal{A} \otimes \mathcal{B}$. So we're left to show

$$\mu \otimes \nu(E) = \int_X \nu(E_x)d\mu = \int_Y \mu(E_y)d\nu \quad (***)$$

Now let $\mathcal{C} = \{E \in \mathcal{A} \otimes \mathcal{B} \text{ s.t. } (***) \text{ holds }\}$. Then
• $A \times B \in \mathcal{C}$ for all $A \in \mathcal{A}, B \in \mathcal{B}$ by definition of $\mu \otimes \nu$,
• $\mathcal{C}$ is stable under complement (because μ, ν are finite),
• $\mathcal{C}$ is stable under countable disjoint unions, and
• $\mathcal{C}$ contains $\{A \times B, A \in \mathcal{A}, B \in \mathcal{B}\}$ which is stable under intersections.

So we may apply the Dynkin Lemma to $\mathcal{C}$ and conclude that $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$.

For the second part where f is only assumed to be integrable, we write $f = f^+ - f^-$ and apply the above to f^+ and f^- separately. $\qquad\square$

Definition. If X is a topological space and μ is Borel measure on X, then μ is called *Radon* if $\mu(K) < \infty \;\forall K$ compact subset of X.

Let $C_c(x)$ be the space of compactly supported functions on X. Clearly every $f \in C_c(x)$ is μ-integrable when μ is Radon.

Indeed $\mu(|f|) \leq \sup_X |f| \cdot \mu(Supp f) < \infty$, where we define $Supp f = \{x \in X; f(x) \neq 0\}$ (support).

So the map $T_\mu C_c(x) \to \mathbb{R}$ by $f \to \mu(f)$ is well defined and *linear*. So T_μ is a linear functional, such that $f \geq 0 \implies T_\mu(f) \geq 0$. We say T_μ is positive.

Theorem. (Riesz representation theorem)
If X is locally compact and second countable (has a countable base), then every positive functional on $C_c(x)$ is of the form T_μ for some (unique) Radon measure μ on X.

2 Probability and Measure

2.1 Introduction

We usually say Ω is a universe of possible outcomes, or the sample space, which is presented as a set by us; and $\mathcal{F}$ is the family of events or observable events, which is presented as a boolean algebra; $\mathbb{P}$ is the probability of an event, which is now equivalent to a measure.

We would like the continuity axiom to be satisfied: if $A_n \in \mathcal{F}$, $A_{n+1} \subseteq A_n$ s.t. $\bigcap_n A_n = \phi$, then $\mathbb{P}(A_n) \xrightarrow{n \to \infty} 0$.

Definition. Let $(X, \mathcal{A}$ be a measurable space. A measure μ on X with $\mu(X) = 1$ is called a probability measure.

Definition. A measurable set $X \to \mathbb{R}$ is called a *random variable*.

Definition. $(X, \mathcal{A}, \mu)$ is called a *probability space* when $\mu(X) = 1$.

In a probability space we usually use $(\Omega, \mathcal{F}, \mathbb{P})$ as standard notation instead of $(X, \mathcal{A}, \mu)$.

Definition. A random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ determines a measure μ_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (where $\mathcal{B}$ denotes the Borel sets) given by $\mu_X(A) = \mathbb{P}(X \in A)$ for all $A \in \mathcal{B}(\mathbb{R})$. We write

$$\mathbb{P}(X \in A) := \mathbb{P}(\{\omega \in \Omega; X(\omega) \subset A\})$$

This measure μ_X is called the *distribution*, or the *law* of X.

Note μ_X is the image of $\mathbb{P}$ under $X : \Omega \to \mathbb{R}$.

We write $X \sim \mu_X$ to mean X is distributed according to μ_X.

Definition. The function $F_X(x) := \mathbb{P}(X \le x)$ $(= \mu_X((-\infty, x]))$ is called the *distribution function* of X.

We replace the integral sign $\int d\mu$ by the expectation sign $\mathbb{E}$, i.e. $\mathbb{E}(X) = \int_\Omega X d\mathbb{P}$, and if $A \in \mathcal{F}$, then $\mathbb{E}(1_A) = \mathbb{P}(A)$.

Example. Toss a coin N times. Let X to be the number of heads, then $\mathbb{P}(X = k) = \binom{N}{k} \frac{1}{2^n}$; this is called the *binomial distribution*. Now $\mathbb{E}(X) = \frac{N}{2} = \sum_1^N k \mathbb{P}(X = k) = \sum_1^n k \binom{N}{k}$; but it's also $\mathbb{E}(\sum_1^N 1_{\varepsilon_k = \text{ heads}}) = \frac{N}{2}$.

Note that F_X is no-decreasing, right continuous, and it determines μ_X. Also, $\lim_{X \to -\infty} F = 0$, and $\lim_{X \to \infty} F = 1$ (*).

Theorem. Conversely given a non-decreasing function $F : \mathbb{R} \to [0, 1]$, which is right continuous and has the above property (*), there exists a unique probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $F = \mu((-\infty, x])$ for all $x \in \mathbb{R}$.

Proof. (sketch)

Define μ on intervals $(a, b]$ as $\mu((a, b]) = F(b) - F(a)$. Extend it to finite union of intervals, and use Caratheodory extension theorem.

$\square$

Definition. If $F_X(x)$ has the form $\int_{-\infty}^{x} f(y)dy$, where dy is a Lebesgue measure, and $f \geq 0$ is Lebesgue integrable, then we say that F_X or X, or μ_X has a *density* w.r.t. Lebesgue measure, and f is called a density function.

Example. Consider a uniform distribution on $[0, 1]$, $f(x) = 1_{[0,1]}$. We have $F(x) \int_{-\infty}^{x} f = 1 - x$.

Example. Consider an exponential distribution with rate $\lambda > 0$: $f_\lambda(x) = \lambda e^{-\lambda x} 1_{x \geq 0}$, $F(x) = 1_{x \geq 0}(1 - e^{-x/\lambda})$.

Example. For a normal distribution (Gaussian), denoted $N(\sigma^2, m)$, we have

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Last time the lecture forgot to say that, μ is called the *Lebesgue-Stieljes measure* associated to F.

Now we'll define independence. Let $\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition. A sequence of events $(A_n)_{n \geq 1}$ is said to be (mutually) independent if for all $I \subseteq \mathbb{N}$ and finite,

$$\mathbb{P}(\bigcap_{i \in I} A_i) = \Pi_{i \in I} \mathbb{P}(A_i)$$

Definition. A sequence of sub-σ-algebras $(\mathcal{A}_n)_{n \geq 1}$ of $\mathcal{F}$ is called *independent* if for all $A_i \in \mathcal{A}_i$, $i \geq 1$, the family $(A_i)_{i \geq 1}$ is independent.

Definition. A sequence $(X_n)_{n \geq 1}$ of random variables is called (mutually) independent if the sequence $(\sigma(X_n))_{n \geq 1}$ is independent.

Notation: If X is a random variable, $\sigma(X) := X^{-1}(\mathcal{B}(\mathbb{R})) = \sigma(\{\omega \in \Omega : X(\omega) \leq x\}_{x \in \mathbb{R}})$ is the smallest σ-algebra making X measurable.

Example. If X takes only finitely many values $x_1, ..., x_k$, then $\sigma(X)$ is a finite boolean algebra whose atoms are $\{\omega, X(\omega) = x_i\}$ for $i = 1, ..., k$.

Example. Suppose $\Omega = \{Heads, Tails\}^2$, i.e. two independent tosses of a coin. Now $\mathbb{P}(X_1 = \varepsilon_1, X_2 = \varepsilon_2) = 1/4$ for all $\varepsilon_1, \varepsilon_2 \in \{H, T\}$.

(missing 1 line)

If $\mathcal{F}$ = discrete boolean algebra on $\Omega = 2^\Omega$. $\sigma(X_1) = \{\{H\} \times \{H, T\}, \{T\} \times \{H, T\}, \phi, \Omega\}$, $\sigma(X_2) = \{\{H, T\} \times \{H\}, \{H, T\} \times \{T\}, \phi, \Omega\}$ are two σ-algebra of events.

Lemma. Let $\mathcal{F}_i \subseteq \mathcal{A}_i$ be a subfamily of events which is stable under finite intersection, s.t. $\sigma(\mathcal{F}_i) = \mathcal{A}_i$. Then to check that $(\mathcal{A}_i)_{i \geq 1}$ forms an independent family, it is enough to check that $\mathbb{P}(\bigcap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i)$ holds for all finite $I \subseteq \mathbb{N}$, and all $A_i \in \mathcal{F}_i$.

Example. a) Suppose $(X_1, ..., X_n)$ are n random variables. Then they are independent iff for all $x_1, ..., x_n \in \mathbb{R}$, $\mathbb{P}(\forall i = 1, ..., n, X_i \leq x_i) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i)$.

Indeed, this is a special case of the lemma setting $\mathcal{F}_i = \{X_i^{-1}((-\infty, x]), x \in \mathbb{R}\}$.

b) If $X_1, ..., X_n$ take only a finite set E of values $(X_1, ..., X_n)$ independent iff for all $e_1, ..., e_n \in E$, $\mathbb{P}(\forall i, X_i = e_i) = \prod_{i=1}^n \mathbb{P}(X_i = e_i)$.

Proof. (of lemma) For two σ-sub algebra $\mathcal{A}_1$ and $\mathcal{A}_2$, let $A_2 \in \mathcal{A}_2$ be s.t. $\mathbb{P}(A \cap A_2) = \mathbb{P}(A)\mathbb{P}(A_2)$ (*) holds for all $A \in \mathcal{F}_1$. Look at the 2 measures $A \to \mathbb{P}(A \cap A_2)$ and $A \to \mathbb{P}(A)\mathbb{P}(A_2)$. The two measure coincide on $\mathcal{F}_1$, hence they coincide on $\mathcal{A}_1$ (by Dynkin's lemma). So (*) holds for all $A \in \mathcal{A}_1$. Now just reverse the roles of $\mathcal{A}_1$ and $\mathcal{A}_2$. $\square$

Example. Let $\Omega = [0, 1]$, $\mathbb{P}$ be a Lebesgue measure, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ be a Borel σ-algebra. Let $X_n(\omega)$ be the nth digit in the decimal expansion of $\omega \in [0, 1]$ (we avoid ambiguity by disallowing ending with 9999..., although it doesn't really matter). We claim that $(X_n)_{n \geq 1}$ is independent. Let $I_{\varepsilon_1...\varepsilon_n} = \{\omega \in \Omega, X_1(\omega) = \varepsilon_1, ..., X_n(\omega) = \varepsilon_n\} = \{\omega \in [0, 1] | [10^n \omega] = \varepsilon_1...\varepsilon_n\}$. We need to check that $\mathbb{P}(X_1(\omega) = \varepsilon_1, ..., X_n(\omega) = \varepsilon_n) = \prod_{i=1}^n \mathbb{P}(X_i(\omega) = \varepsilon_i)$, but it's clear because $\mathbb{P}(\omega \in I_{\varepsilon_1...\varepsilon_n}) = 1/10^n$ and RHS is $1/10^n$ as well.

Remark. If $f_1, ..., f_n$ are Borel measurable functions $|R \to \mathbb{R}$ and $X_1, ..., X_n$ are independent random variables, then so are $f_1(X_1), ..., f_n(X_n)$. The reason is that $\sigma(f(X)) \subseteq \sigma(X)$.

Remark. Independence implies pairwise independence, but not conversely. Bernstein's example: Let X, Y be two independent coin tosses, $X = 1$ for heads, $X = 0$ for tails and same for Y. Now set $Z = |X - Y|$. We can check that each pair among X, Y, Z are independent, but clearly Z cannot independently take value from X and Y.

Proposition. Let $X_1, ..., X_n$ be n random variables. $(X_1, ..., X_n)$ are mutually independent if and only if $\mu_{X_1,...,X_n} = \mu_{X_1} \otimes ... \otimes \mu_{X_n}$ (product measures). Here $\mu_{(X_1,...,X_n)}$ is the *law* or *distribution* of $(X_1, ..., X_n)$, sometimes also called the *joint law* or *joint distribution*.

It's a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ defined by $\mu_{(X_1,...,X_n)}(A) = \mathbb{P}((X_1, ..., X_n) \in A)$ for all $A \in \mathcal{B}(\mathbb{R}^n)$.

Proof. $\mu_{(X_1,...,X_n)}(\prod_{i=1}^n (-\infty, x_i)) = \mathbb{P}((X_1, ..., X_n) \in \prod_1^n (-\infty, x_i]) = \mathbb{P}(\forall i, X_i \leq x_i)$. So independence if and only if $\forall x_i \in \mathbb{R}, \mathbb{P}(\forall i, X_i \leq x) = \prod_1^n \mathbb{P}(X_i \leq x_i)$, if and only if $\mu_{(X_1,...,X_n)}(\prod_1^n (-\infty, x_i]) = \prod \mu_{X_i}((-\infty, x_i])$. $\square$

Recall from last time, that A, B events are called independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Proposition. Suppose X and Y are two independent random variables, and X, Y are integrable ($\mathbb{E}(|X|) < \infty, \mathbb{E}(|Y|) < \infty$). Then XY is integrable, and $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ (*).
Moreover, (*) holds also if X and Y are just assumed positive.

Proof. Recall that, X, Y are independent iff the joint law $\mu_{(x,y)}$ is the product $\mu_x o \times \mu_y$, where μ_x, μ_y are the law of X and Y respectively. So we can apply Fubini:

$$\mathbb{E}(X) = \int_{\mathbb{R}} x d\mu_x, \mathbb{E}(Y) = \int_{\mathbb{R}} y d\mu_y, \mathbb{E}(XY) = \int_{\mathbb{R} \times \mathbb{R}} xy d\mu_{(x,y)}$$

$\square$

Proposition. (1st Borel-Cantelli lemma)
Let $(\Omega, \mathcal{F}.\mathbb{R})$ be a probability space. Let $(A_n)_{n \geq 1}$ be a sequence of events, and assume $\sum_{n \geq 1} \mathbb{P}(A_n) < \infty$. Then $\mathbb{P}(\limsup A_n) = 0$.

Proof.

$$\limsup A_n := \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n$$

so

$$\mathbb{P}(\limsup A_n) \leq \mathbb{P}(\bigcup_{n \geq m} A_n) \leq \sum_{n \geq m} \mathbb{P}(A_n) \xrightarrow{m \to +\infty} 0$$

$\square$

Example. Take $\Omega = [0, 1]$, $\mathbb{P}$ be the Lebesgue measure, $\mathcal{F}$ be the Borel σ-algebra. Define a number $\alpha \in \mathbb{R}$ to be *very-well-approximated (VWA)* if $\exists \varepsilon > 0$ there are infinitely many integers p, q s.t. $|\alpha - p/q| \leq \frac{1}{q^{2+\varepsilon}}$. We claim that almost every number is not VWA (this is an exercise; just apply Borel-Cantelli to the events $A_q = \{x \in [0, 1], d(qx, \mathbb{Z}) < \frac{1}{q^{1+\varepsilon}}\}$.

Proposition. (2nd Borel-Cantelli lemma, converse)
Assume that the events $(A_n)_n$ are independent, and $\sum_{n \geq 1} \mathbb{P}(A_n) = +\infty$. Then $\mathbb{P}(\limsup A_n) = 1$.

Remark. Conclusion fails without the independence assumption. For example, $A_n = [0, \frac{1}{n}] \subseteq \Omega = [0, 1]$, $\sigma\mathbb{P}(A_n) = +\infty$ but $\limsup A_n = \{0\}$.

Proof. $\limsup A_n = \bigcap_{m \geq 1} \bigcup_{n \geq m} A_m$, so $(\limsup A_n)^C = \bigcup_{m \geq 1} \bigcap_{n \geq m} A_m^C$. But for all m,

$$\mathbb{P}(\bigcap_{n \geq m} A_m^C) = \prod_{n \geq m} \mathbb{P}(A_n^C)$$
$$= \prod_{n \geq m} (1 - \mathbb{P}(A_n))$$
$$\leq \prod_{n \geq m} e^{-\mathbb{P}(A_n)}$$
$$= e^{-\sum_{n \geq m} \mathbb{P}(A_n)}$$
$$= e^{-\infty} = 0$$

$\square$

Definition. A random process is an infinite family $(X_n)_{n \geq 1}$ and real random variable.

Let n be the time parameter, $\mathcal{F}_n$ (time filtration) be $\sigma(X_1, ..., X_n)$ = smallest sub-σ-algebra of $\mathcal{F}$ which makes all $X_i, i \leq n$ measurable, so $\mathcal{F}_{n+1} \supseteq \mathcal{F}_n$.

Definition. The tail σ-algebra of $(X_n)_{n \geq 1}$ is defined as

$$\mathcal{T} = \bigcap_{n \geq 1} \sigma(X_n, X_{n+1}, ...)$$

Example. The following events belong to $\mathcal{T}$:

$$\{\limsup X_n \geq T\} \, or \, \{\omega \in \Omega | (X_n(\omega))_{n \geq 1} \text{ converges}\}$$

Definition. (Kolmogorov 0-1 law)
Suppose $(X_n)_{n \geq 1}$ is a sequence of independent random variable. Then the tail σ-algebra $\mathcal{T}$ is trivial, i.e. $\forall A \in \mathcal{T}$, $\mathbb{P}(A) \in \{0, 1\}$.

Proof. Let $A \in \mathcal{T}$, let $B \in \sigma(X_1, ..., X_n)$. Then A and B are independent, because $\sigma(X_1, ..., X_n)$ and $\mathcal{T}$ are independent. So $\mathbb{P}(A \bigcap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. Now the measures $B \to \mathbb{P}(A \bigcap B)$ and $B \to \mathbb{P}(A)\mathbb{P}(B)$ coincide on $\sigma(X_1, ..., X_n)$ for all n; so they coincide on $\sigma(X_1, ..., X_n, ...)$. But $\mathcal{T} \subseteq \sigma(X_1, ..., X_n, ...)$. In particular, $\mathbb{P}(A \bigcap A) = \mathbb{P}(A)\mathbb{P}(A)$. $\square$

Example. Take an iid sequence $(X_n)_{n \geq 1}$ of RV at $T \geq 0$, $\mathbb{R}(X_1 < t) < 1$. Then as $\limsup X_n = +\infty$. Indeed, $\sum_{k=1}^{n} \mathbb{P}(X_k \geq T) = n\mathbb{P}(X_1 \geq T) = n(1 - \mathbb{P}(X_1 < T)) \xrightarrow{n \to \infty} \infty$, so the 2nd Borel-Cantelli lemma applies and $\mathbb{P}(\limsup A_n) = 1$.

Example. Take $(\varepsilon_n)_{n \geq 1}$ be iid RV, with $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = 1/2$. Take $(a_n)_{n \geq 1}$ some sequence of real numbers. When does $\sum_{n \geq 1} \varepsilon_n a_n$ converge?

The Kolmogorov 0-1 law tells us that this happens with probability 0 or 1 depending only on the sequence $(a_n)_{n \geq 1}$.

Theorem. (Rademochan-Paley-Zygmund)
$\sum \varepsilon_n a_n$ converges almost surely iff $\sum a_n^2 < \infty$.

Let μ be a probability measure on $\mathbb{R}$. Can you find a probability space $(\Omega, \mathcal{F}.\mathbb{P})$ and a RV X s.t. μ is the law of X? Yes: $\Omega = \mathbb{R}$, $\mathcal{F}$ be the Borel sets, $\mathbb{P} = \mu$, and $X(\omega) = \omega$.

Definition. If X is a RV, the kth moment of X is, by definition, $\mathbb{E}(X^k)$ for $k \in \mathbb{N}$.

Some additional basic terminologies: for X a random variable, $\mathbb{E}(X)$ is called the *mean*, $\mathbb{E}((X - \mathbb{E}X)^2)$ is called the variance. We have $Var X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$. Also if X, Y are independent then $Var(X + Y) = Var X + Var Y$.

Theorem. (Strong law of large numbers (under a 4th moment assumption))
Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed (i.i.d.) random variables with common law μ. Assume $\int_{\mathbb{R}} x^4 d\mu(x) < \infty$ ($\iff$ $\mathbb{E}(|X_1|^4 < \infty)$, then

$$\frac{1}{n}\sum_{k=1}^{n} x_k \xrightarrow{a.s.} \mathbb{E}(X_1)$$

Example. $\Omega = [0,1]$, $\mathbb{P} = \mathcal{L}$, $\omega = 0.\varepsilon_1\varepsilon_2...\varepsilon_n...$, $\varepsilon_i \in \{0,...,9\}$ decimal expansion of ω.

We observe that $(\varepsilon_n(\omega))_{n\geq 1}$ were independent and uniformly distributed in $\{0,1,...,9\}$. Set $X_n^{(i)}(\omega) = 1_{\{\varepsilon_n(\omega)=i\}}$. These are i.i.d. random variables and $|X_n^{(i)}| \leq 1$ ($\implies$ 4th moment $< \infty$). So the SSC applies that

$$\frac{1}{n}\sum_{k=1}^{n} X_k^{(i)} = \frac{1}{n}\#\{k \leq n, \varepsilon_k(\omega) = i\} \to \mathbb{E}(X_1^{(i)}) = \frac{1}{10}$$

for Lebesgue a.e., i.e. ω the proportion of each digit is the same (normal number).

Cauchy Schwartz inequality: let f,g be square integrable functions on $(X,\mathcal{A},\mu)$. Then fg is integrable, and

$$\left|\int fg d\mu\right| \leq \sqrt{\int f^2 d\mu}\sqrt{\int g^2 d\mu}$$

In probabilistic terms, this means: if X,Y are r.v. s.t. $\mathbb{E}(X^2) < \infty, \mathbb{E}(Y^2) < \infty$, then $\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}X^2 \cdot \mathbb{E}Y^2}$.

Proof. Look at $t \to \mathbb{E}((X + tY)^2) = \mathbb{E}X^2 + t^2\mathbb{E}Y^2 + 2t\mathbb{E}(XY)$ which is non-negative for all t. Then consider discriminant. $\square$

Proof. (of SLL)
wlog we may assume that $\mathbb{E}(X_1) = 0$, because replace X_n by $X_n - \mathbb{E}(X_1)$. This is allowed because $\mathbb{E}((X_n - \mathbb{E}X_1)^4) < \infty$, indeed

$$\begin{aligned}(X_n - \mathbb{E}X_1)^4 &= [(X_n - \mathbb{E}X_1)^2]^2 \\ &\leq (2X_n^2 + 2(\mathbb{E}X_1)^2)^2 \\ &\leq 8[X_n^4 + (\mathbb{E}X_1)^4]\end{aligned}$$

So we have $\mathbb{E}(X_n) = 0$ for all n. Now let $S_n = \frac{1}{n}(X_1 + ... + X_n)$. Then after some calculation we get

$$\begin{aligned}\mathbb{E}(S_n^4) &= \frac{1}{n^4}(X_1^4 + ... + X_n^4 + 6\sum_{i<j}\mathbb{E}X_i^2 \cdot \mathbb{E}X_j^2) \\ &= \frac{1}{n^4}\left(n\mathbb{E}(X_1^4) + 6\frac{n(n-1)}{2}(\mathbb{E}(X_1^2))^2\right) \\ &= O(\frac{1}{n^2})\end{aligned}$$

So we conclude that

$$\sum_{n \geq 1} \mathbb{E}(S_n^4) = \mathbb{E}(\sum_{n \geq 1} S_n^4) < \infty$$

So $\sum_{n \geq 1} S_n^4 < \infty$ a.s. $\implies \lim n \to \infty S_n = 0$ a.s.. $\square$

Now we discuss the types of convergence of r.v.'s:

Definition. A sequence of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ say $(\mu_n)_{n \geq 1}$ is said to *converge weakly* to a measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ if for all $f \in C_b(\mathbb{R}^d), \mu_n(f) \xrightarrow{n \to \infty} \mu(f)$. We denote $C_b(\mathbb{R}^d) =$ space of continuous and bounde real functions on $\mathbb{R}^d$.

Example. We take $\mu_n = \delta_{x_n}$ a Dirac mass (delta functions) with $x_n \in \mathbb{R}^d$ with $x_n \to x \in \mathbb{R}^d$. Then clearly $\mu_n \to \delta_x$ weakly.

Example. $\mu = N(0, \sigma_n^2$ (=centered gaussian with standard deviation σ_n). Assume $\sigma_n \to 0$, then $\mu_n \to \delta_0$ weakly.

Definition. A sequence of $\mathbb{R}^d$-valued random variables $(X_n)_{n \geq 1}$ is said to converge to X
(a) almost surely (a.s.), if for $\mathbb{P}-$a.e ω, $X_n(\omega) \xrightarrow{n \to \infty} X(\omega)$;
(b) in probability (or in measure), if $\forall \varepsilon > 0, \mathbb{P}(||X_n - X|| > \varepsilon) \xrightarrow{n \to \infty} 0$;
(c) in distribution, if $\mu_{X_n} \xrightarrow{n \to \infty} \mu_X$ weakly (μ_{X_n} is the law of X_n on $\mathbb{R}^d =$ probability on $\mathbb{R}^d$).

Remark. We have (a) implies (b), (b) implies (c), but not the other way in general.
(a) $\to$ (b): $\mathbb{P}(||X_n - X|| > \varepsilon) = \mathbb{E}(1_{||X_n - X|| > \varepsilon}) \xrightarrow{n \to \infty} 0$ by LDC.
(b) $\to$ (c): Let $f \in C_b(\mathbb{R}^d)$. f is uniformly continuous on compact subsets of $\mathbb{R}^d$. So $\forall \varepsilon > 0, \exists \delta > 0$ s.t. if $||x|| < 1/\varepsilon$, $||x - y|| < \delta$, then $|f(x) - f(y)| < \varepsilon$. So

$$
\begin{aligned}
|\mu_n(f) - \mu(f)| &\leq |E(|f(X_n) - f(X)| \\
&\leq \mathbb{E} \left(1_{||X_n - X|| < \delta} 1_{||X|| < 1/\varepsilon} |f(X_n) - f(X)| \right) + 2||f||_\infty \mathbb{P}(||X_n - X|| > \delta \text{ or } ||X|| > 1/\varepsilon) \\
&\leq \varepsilon + 2||f||_\infty [\mathbb{P}(||X_n - X|| > \delta) + \mathbb{P}(||X|| > 1/\varepsilon)]
\end{aligned}
$$

So

$$\limsup_n |\mu_n(f) - \mu(f)| \leq \varepsilon + 2||f||_\infty \mathbb{P}(||X|| > 1/\varepsilon) \xrightarrow{\varepsilon \to 0} 0$$

Remark. When $d = 1$, (c) is equivalent to saying that the distribution function $F_{\mu_n}(X) \to F_\mu(x)$ for all $x \in \mathbb{R}$ where $F(x)$ is continuous. (exercise). Here $F_\mu(x) := \mu((-\infty, x])$, $F_{\mu_n}(x) := \mu_n((-\infty, x])$.
To see why continuity is useful, e.g. $\mu_n = \delta_{1/n} \to \delta_0 = \mu$, $F_{\mu_n} = 1_{x \geq \frac{1}{n}}$, $F_\mu = 1_{x \geq 0}$. $F_{\mu_n}(0) \nrightarrow F_\mu(0)$.

Proposition. If $(X_n)_{n \geq 1}$ converges in probability to X, then $\exists$ subsequence $(X_{n_k})_{k \geq 1}$ s.t. $X_{n_k} \xrightarrow{k \to \infty} X$ a.s..

Proof. for all k we have $\mathbb{P}(||X_n - X|| > 1/k) \xrightarrow{n \to \infty} 0 \implies \exists n_k$ s.t. $\mathbb{P}(||X_n - X|| > 1/k) \leq \frac{1}{2^k}$. So $\sum \mathbb{P}(||X - n - X|| > 1/k) < \infty \implies \mathbb{P}(\limsup ||X_{n_k} - X|| > 1/k) = 0$ by Borel-Cantelli. $\square$

Yet we introduce another type of convergence:

Definition. (convergence in mean, or L^1)
A sequence of $\mathbb{R}^d$-valued integrable random variable $(X_n)_{n\geq 1}$ converges in L^1 or in mean towards an integrable r.v. X if

$$\mathbb{E}(||X_n - X||) \xrightarrow{n\to\infty} 0$$

which we'll write $X_n \xrightarrow{L^1} X$.

Remark. • if $X_n \xrightarrow{L^1} X$ then $X_n \to X$ in probability. Indeed, for all $\varepsilon > 0$,

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

• The markov inequality,

$$\varepsilon\mathbb{P}(|X| \geq \varepsilon) \leq \mathbb{E}(|X|)$$

and the Chebyshev inequality,

$$\varepsilon^2\mathbb{P}(|X - \mathbb{E}X| > \varepsilon) \leq VarX$$

hold for any r.v. X and for any $\varepsilon > 0$.

The converse does not hold in general (convergence in probability doesn't imply convergence in L^1: for example, take $\Omega = [0, 1]$, $\mathbb{P}$=Lebesgue measure, $X_n = n1_{[0,1/n]}$, but $\mathbb{E}(X_n = 1$ for all n.

• If $\exists M \geq 0$ s.t. $||X_n|| \leq M$ a.s. for all n, then the converse holds:

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

So

$$\limsup \mathbb{E}(||X_n - X||) \leq \varepsilon$$

Definition. A sequence of integrable random variables $(X_n)_{n\geq 1}$ is called uniformly integrable (or U.I.) if

$$\sup_{n\geq 1} \mathbb{E}(||X_n||1_{||X_n||>M}) \xrightarrow{M\to\infty} 0$$

Remark. If $X_n = X$ for all n, and X is integrable, then $(X_n)_n$ is U.I.:

$$\mathbb{E}(||X||1_{||X||\leq M}) \uparrow \mathbb{E}(||X||)$$

as $M \to \infty$ by MCT, so

$$\mathbb{E}(||X||1_{||X||>M}) \xrightarrow{M\to\infty} 0$$

Suppose the X_n's are dominated, i.e. there exists Y integrable r.v. s.t. $||X_n|| \leq Y$ for all n. Then $(X_n)_n$ are U.I.:

$$\mathbb{E}(||X_n||1_{||X_n||>M}) \leq \mathbb{E}(Y1_{Y>M}) \xrightarrow{M\to\infty} 0$$

If $X_n \xrightarrow{L^1} X$, then $(X_n)_n$ is U.I.:

$$\mathbb{E}(||X_n||1_{||X_n||>M}) \leq \mathbb{E}(||X_n - X||1_{||X_n||>M}) + \mathbb{E}(||X||1_{||X_n||>M})$$
$$\leq \mathbb{E}(||X_n - X||) + \mathbb{E}(||X||1_{||X||>M}1_{||X_m||>M}) + \mathbb{E}(||X||1_{||X||<M}1_{||X_n||>M})$$
$$\leq \mathbb{E}(||X_n - X||) + \mathbb{E}(||X||1_{||X||>M}) + M\mathbb{E}(1_{||X_n||\geq M})$$
$$\leq \mathbb{E}(||X_n - X||) + \varepsilon(M)$$

So $M\mathbb{P}(||X||_n > M) \leq \mathbb{E}(||X_n||) \to \mathbb{E}(||X||)$ (?)

Theorem. Let $(X_n)_{n\geq 1}$ be a sequence of integrable r.v.'s. Let X be another random variable. $(X_n)_n$ is U.I. and converges in probability to X $\iff$ X is integrable and $X_n \xrightarrow{L^1} X$.

Proof. Suppose $(X_n)_n$ is UI and converges to X in probability. Last time we've shown that $\exists (X_{n_k})_k$ of $X_{n_k} \xrightarrow{k\to\infty} X$ a.s., and

$$\mathbb{E}(||X||1_{||X||>M}) \leq \underbrace{\liminf_{k\to\infty} \mathbb{E}(||X_{n_k}||1_{||X_{n_k}>M})}_{\varepsilon(M)}$$

by Fatou's lemma. So

$$\mathbb{E}(||X||) = \mathbb{E}(||X||1_{||X||<M}) + \mathbb{E}(||X||1_{||X||\geq M}) \leq M + \varepsilon(M) < \infty$$

so X is integrable. Hence

$$\mathbb{E}(||X_n - X||) = \underbrace{\mathbb{E}(1_{||X_n-X||>\varepsilon}||X_n - X||)}_{(*)} + \underbrace{\mathbb{E}(1_{||X_n-X||\leq\varepsilon}||X_n - X||)}_{\leq\varepsilon}$$

We have

$$(*) = \mathbb{E}(1_{||X_n-X||>\varepsilon}(1_{||X_n||<M}1_{||X||<M} + 1_{||X_n||\geq M}1_{||X||<M} + 1_{||X_n||\geq M}1_{||X||\geq M})||X_n - X||)$$
$$\leq 2M\mathbb{P}(||X_n - X|| > \varepsilon) + \mathbb{E}(1_{||X_n-X||>\varepsilon}(M + ||X_n||)1_{||X_n||>M})$$
$$+ \mathbb{E}(1_{||X_n-X||>\varepsilon}(M + ||X||)1_{||X||>M}) + \mathbb{E}(||X_n||1_{||X_n\geq M} + ||X||1_{||X||>M})$$
$$\leq 4M\mathbb{P}(||X_n - X|| > \varepsilon) + 2\underbrace{\mathbb{E}(||X_n||1_{||X_n||>M})}_{\varepsilon(M)} + 2\underbrace{\mathbb{E}(||X||1_{||X||>M})}_{\varepsilon(M)}$$

So

$$\limsup \mathbb{E}(||X_n - X||) \leq 4\varepsilon(M)$$

Let $M \to \infty$, $\varepsilon \to 0$. $\qquad\qquad\square$

Remark. This theorem subsumes the Lebesgue Dominated Converged Theorem.

If $\exists C > 0$ and some $p > 1$ s.t. $\mathbb{E}(||X_n||^p) \leq C$ for all n, then $(X_n)_n$ is UI:

$$M^{p-1}\mathbb{E}(||X_n||1_{||X_n||>M}) \leq \mathbb{E}(||X_n||^p) \leq C$$

—2017-11-13—

Proof. (of the theorem that we failed to proof last time)

• Backward: need to show if $X_n \xrightarrow{L^1} X$ then $(X_n)_n$ is UI.

Remark. $(X_n)_n$ is UI iff $\limsup_{n\to\infty} \mathbb{E}(||X_n||1_{||X_n||>M}) \xrightarrow{M\to\infty} 0$. Reason is: for any finitely many $X_1, ..., X_m$, $\mathbb{E}(||X_i||1_{||X_i||>M}) \xrightarrow{M\to\infty} 0$ for each $i = 1, ..., n$.

$$\begin{aligned}
\mathbb{E}(||X_n||1_{||X_n||>M}) &\leq \mathbb{E}(||X_n - X||) + \mathbb{E}(||X||1_{||X_n||>M}) \\
&\leq \mathbb{E}(||X_n - X||(+\mathbb{E}(||X||1_{||X_n||>M}(1_{||X||<M/2} + 1_{||X||\geq M/2})) \\
&\leq \mathbb{E}(||X_n - X||) + \mathbb{E}(||X||(1_{||X_n-X||>M/2}1_{||X||<M/2} + 1_{||X||>M/2})) \\
&\leq \mathbb{E}(||X_n - X||)(\to 0) + \frac{M}{2}\mathbb{P}(||X_n - X|| > M/2)(\to 0) + \mathbb{E}(||X||1_{||X||>M/2})
\end{aligned}$$

so

$$\limsup \mathbb{E}(||X_n||1_{||X_n||>M}) \leq \mathbb{E}(||X||1_{||X||>M/2}) \xrightarrow{M\to\infty} 0$$

$\square$

Remark. $(X_n)_n$ is UI iff $\sup_n \mathbb{E}(||X_n||) < \infty$ and $\forall \varepsilon > 0 \exists \delta > 0 \forall$ event $A, \mathbb{P}(A) < \delta \implies \sup_n \mathbb{E}(||X_n||1_A) < \varepsilon$.

Proof. Forward: $\forall \varepsilon > 0 \exists M_0(\varepsilon) \forall M \geq M_0$, $\mathbb{E}(||X_n||1_{||X_n||>M}) < \varepsilon$ by definition of UI.
St $\delta = \varepsilon/M_0(\varepsilon)$, so

$$\begin{aligned}
\mathbb{E}(||X_n||1_A) &\leq \mathbb{E}(||X_n||1_A 1_{||X_n||>M_0}) + \mathbb{E}(||X_n||1_A 1_{||X_n||\leq M}) \\
&\leq \varepsilon + M_0\mathbb{P}(A) \leq 2\varepsilon
\end{aligned}$$

Backward: $M\mathbb{P}(||X_n|| > M) \leq \mathbb{E}(||X_n||) \leq \sup_n \mathbb{E}(||X_n||) < \infty \implies \mathbb{P}(||X_n|| \geq M) \xrightarrow{M\to\infty} 0$. $\square$

We'll move on to L^p spaces.

Proposition. (Holder's inequality)
Let $p, q \in [1, +\infty]$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$, with the convention that $\frac{1}{\infty} = 0$. Let f, g be measurable functions on a measurable space $(X, \mathcal{A}, \mu)$. Then

$$\int_X |fg|d\mu \leq p\sqrt{\int_X |f|^p d\mu} \cdot q\sqrt{\int_X |g|^q d\mu}$$

with equality iff $\exists a, b \in \mathbb{R}$ not both 0 s.t. $a|f|^p = b|g|^q$ μ-a.e..

Proof. WLOT we can assume $p, q \in (1, \infty)$. Also we can assume that $\int |f|^p d\mu = 1 = \int |g|^q d\mu$ (up to changing f and g by a scalar).

Young's inequality for products: $\forall a, b \geq 0$, $ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q$ (this follows from convexity of $x \to -\log x$: $\log(ta^p + (1-t)b^q) \leq t\log a^p + (1-t)\log b^q$, for

$t \in [0,1]$. Then take $t = \frac{1}{p}$). Write $|f(x)g(x)| \leq \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q$, and integrate we get $\int |fg|d\mu \leq 1$.

But $-\log$ is trictly conves, so equality holds only if we have equality a.e. that $|f(x)|^p = |g(x)|^q$. $\qquad\square$

Proposition. (Minkowski inequality)
Let $p \in [1, \infty]$, f, g measurable on a measure space $(X, \mathcal{A}, \mu)$. Then

$$\sqrt[p]{\int |f+g|^p d\mu} \leq \sqrt[p]{\int |f|^p d\mu} + \sqrt[p]{\int |g|^p d\mu}$$

Proof. For $p = 1$ this is obvious by linearity of $\int$. For $p > 1$, apply Holder's to $|f||f+g|^{p-1}$ and to $|g||f+g|^{p-1}$, we get

$$\int |f+g|^p \leq \int (|f|+|g|)|f+g|^{p-1} \leq \left(\sqrt[p]{\int |f|^p} + \sqrt[p]{\int |g|^p} \right) \sqrt[q]{\int |f+g|^{q(p-1)(=p)}}$$

$\qquad\square$

We need to check that $\int |f+g|^p < \infty$. This follows from $|f+g|^p \leq (2 \max |f|, |g|)^p = 2^p \max |f|^p, |g|^p \leq 2^p(|f|^p + |g|^p)$. So integrability of $|f|^p$ and $|g|^p$ implies the result.

Notation. Let $\mathcal{L}^p(X, \mathcal{A}, \mu)$ be the set of functions that are $\mathcal{A}$-measurable of X s.t. $\int_X |f|^p d\mu < \infty$. By the previous remark, this is a real vector space.

Definition. The L^p-norm of f is defined by

$$||f||_p = \sqrt[p]{\int |f|^p d\mu}$$

for $1 \leq p < \infty$. The following holds:
• $||\alpha f||_p = |\alpha| ||f||_p \forall \alpha \in \mathbb{R}$;
• $||f + g||p \leq ||f||_p + ||g||_p$ (Minkowski' inequality implies subadditvity).

Remark. If $||f||_p = 0$, then $f = 0$ μ-a.e.. So $||\cdot||_p$ is only a pseudo-norm on $\mathcal{L}^p(X, \mathcal{A}, \mu)$.

Note: if $f_1 = f_2$ μ-a.e., $g_1 = g_2$ μ-a.e., then $f_1 + g_1 = f_2 + g_2$ μ-a.e., $f_1 g_1 = f_2 g_2$ μ−a.e.. So if we say that two functions f, g on $(X, \mathcal{A}, \mu)$ are equivalent, write $f \overset{\mu}{\sim} g$ if $f - g = 0$ μ-a.e., then this is an equivalence relation compatible with $+$ and $\times$.

Definition. Let $L^p(X, \mathcal{A}, \mu)$ be the set of equivalence classes of functions in $\mathcal{L}^p(X, \mathcal{A}, \mu)$ (note that James Norris chose a different notaiton for this one). Define $||[f]||_p = ||f||_p$. This is a real vector space and $||[]||_p$ is a genuine norm on it, where $[f]$ if the equivalence class of f.

Definition. When $p = +\infty$, we define $||f||_\infty = \inf\{t \geq 0, |f(x)| \leq t \text{ for } \mu\text{-a.e.} x\}$.

Note that this depends on the measure, and is not just a simple maximum of f.

Proposition. (completeness) $L^p(X, \mathcal{A}, \mu)$, $p \in [1, \infty]$ is complete, i.e. every Cauchy sequence converges, iff If $(f_n)_{n\geq 1}$, $f_n \in \mathcal{L}^p(X, \mathcal{A}, \mu)$ and $\forall \varepsilon \exists N ||f_n - f_m||_p < \varepsilon \forall n, m > N$, then $\exists f \in \mathcal{L}^p(x, \mathcal{A}, \mu)$ s.t. $||f_n - f||_p \xrightarrow{n \to \infty} 0$.

Proof. Pick a subsequence n_k s.t. $\sum_{k \geq 1} ||f_{n_{k+1}} - f_{n_k}|| \leq \varepsilon$.

By Minkowski's inequality, $||\sum_{k=1}^{K} |f_{n_{k+1}} - f_{n_k}|||_p \leq \varepsilon$ for all $K < \infty$. Let $K \to \infty$, so by MCT, it holds also for $K = \infty$. $\implies$ μ-a.e. $\sum_1^\infty |f_{n_{k+1}} - f_{n_k}|(x) < \infty \implies$ for μ-a.e. x, $(f_{n_k}(x))_k$ converges (by completeness of $\mathbb{R}$), $\implies$ let $f(x)$ be the limit, outside the set of x's, just let $f(x) = 0$. Finally, $||f_n - f||_p = |\lim_{k \to \infty}(f_n - f_{n_k})||_p \leq \liminf_{k \to \infty} ||f_n - f_{n_k}||_q$ by Fatou's lemma. But since f_n is Cauchy, the above is $\leq \varepsilon$ if n is large enough. $\square$

Last time we defined a $\mathcal{L}^p(X, \mathcal{A}, \mu)$ space for $p \in (1, \infty)$, as well as $\mathcal{L}^\infty$. If $\mathcal{L}^p$ is a complete normed vector space, we call it a *Banach space*.

Remark. (approximation by simple functions)
The vector space of simple functions of the form $\sum_{i=1}^{N} a_i 1_{A_i}$ where $a_i \in \mathbb{R}$, $A_i \in \mathcal{A}_i$ s.t. $\mu(A_i) < \infty$ *is dense* in $L^p(X, \mathcal{A}, \mu)$ for all $p \in [1, \infty]$. Namely, if $f \in L^p(X, \mathcal{A}, \mu)$ then let $f = f^+ - f^-$ ($f_n^+ = \max([n^2 f^+/n^2], n)$ is a simple function), and $f_n^+ \to f^+$, $f_n^- \to f^-$ a.e., and $|f_n| \leq 2|f|$. So when $p < \infty$ by DCT we get $||f_n - f||_p \xrightarrow{n \to \infty} 0$, and when $p = \infty$, $||f - f_n|| \leq \frac{1}{n^2}$.

In particular, $L^\infty \cap L^p$ (which is simple finite) is dense in L^p.

Remark. If $X = \mathbb{R}^d$, $\mathcal{A} =$ Borel σ-salgebra, $\mu =$ Lebesgue measure, then $C_c(\mathbb{R}^d)$(continuous functions with compact support) and $C_c^\infty(\mathbb{R}^d)$ (smooth functions with compact support are dense in $L^p(\mathbb{R}^d, Leb)$ when $p \in [1, \infty)$. However, this is not true for $p = \infty$. For example, $||1_{[0,1]} - f||_\infty \geq 1/2$ if f is continuous.

Note that it's possible to have a smooth function with compact support: consider $f(x) = 0$ if $x \leq 0$ and $f(x) = e^{-1/2}$ if $x > 0$, which is in $C^\infty(\mathbb{R})$. Then let $f_2 = f(x)f(1-x)$, and $\phi(x) = \int_{-\infty}^{x} f_2(u)du$ we get our desired function.

For $I = [a, b]$, $\forall \varepsilon$, $\exists \psi_{I,\varepsilon} \in C^\infty(\mathbb{R})$ with $1_{[a,b]} \leq \psi_{I,\varepsilon} \leq 1_{a-\varepsilon, b+\varepsilon}$. In $\mathbb{R}^d$, $B = \prod_i^d [a_i, b_i]$ a box, $B_\varepsilon = \prod_1^d [a_i - \varepsilon, b_i + \varepsilon]$, define $\psi_{B,\varepsilon}(x) = \prod_1^d \psi_{[a_i, b_i], \varepsilon}(x_i)$ $x = (x_1, ..., x_l$, and then $\psi_{B,\varepsilon} \in C_c^\infty(\mathbb{R}^d)$, $1_B \leq \psi_{B,\varepsilon} \leq 1_{B_\varepsilon}$.

Now if A is a bounded Borel set, there exists K compact, U open, $K \subseteq A \subseteq U$, $\mu(U \setminus K) < \varepsilon$. Using a cover of K by boxes contained in U, we can build a function $\psi \in C_c^{(}\mathbb{R}^d)$ s.t. $1_K \leq \psi \leq 1_U$.

2.2   Hilbert spaces

A *Hilbert space* is a vector space (over $\mathbb{R}$ or over $\mathbb{C}$), call it H, which is endowed with an *inner product*, i.e. a map $H \times H \to \mathbb{R} or \mathbb{C} : (x, y) \to <x, y>$, s.t.
- $<x, x> \in \mathbb{R}^+$, $<x, x> = 0 \implies x = 0$;
- $<x, y> = <y, x>$ for all $x, y \in H$, and
- $<x + \lambda x', y> = <x, y> + \lambda <x', y>$ for all $\lambda \in \mathbb{R}$ ($or \mathbb{C}$), and
- H is complete w.r.t. the norm $||x||_i = \sqrt{<x, x>}$.

For example, $H = \mathbb{R}^d$ with $<x, y> = \sum_1^d x_i y_i$ or $H = \mathbb{C}^d$ with $<x, y> = \sum_1^d x_i \bar{y}_i$ are both Hilbert.

Another example is, if $(X, \mathcal{A}, \mu)$ is a measure space, $H = L^2(X, \mathcal{A}, \mu)$, and

$$<f, g> = \int_X fg d\mu$$

which is well defined by Cauchy-Schwartz. The H is a real Hilbert space.

Lemma. (existence of orthogonal projection)
Let H be a Hilbert space, and let V be a closed vector subspace. Then for all $x \in H$, $\exists! y \in V$ s.t. $d(x, V) = ||x - y||$ (here $d(x, V) = \inf\{||x - z||, z \in V\}$).

Proof. Pick a sequence $y_n \in V$ s.t. $||x - y_n|| \to d(x, V)$. Note that $||y_n - y_m|| \leq ||x - y_n|| + ||x - y_m||$. So $(y_n)_n$ is a Cauchy sequence. H complete implies that $y_n \to y \exists y \in H$, V closed implies $y \in V$ which shows the existence of y. To show uniqueness, if y, y' both satisfy the requirements. Then $||x - \frac{y+y'}{2}||^2 + \frac{1}{4}||y - y'||^2 = \frac{1}{2}(||x - y||^2 = ||x - y'||^2)$. So $||y - y'|| = 0$ as the first term in LHS and both terms in RHS are $d(x, V)^2$, i.e. $y = y'$.

Oops, but $(y_n)_n$ is not a Cauchy sequence as $||x - y_n|| \not\to 0$. So this proof doesn't work. $\square$

Remark. The parallelogram identity holds in any Hilbert space: $\forall a, b \in H$, $||a+b||^2 + ||a-b||^2 = 2(||a||^2 + ||b||^2)$. This is because $||a+b||^2 = <a+b, a+b> = <a,> + <a, b> + <b, a> + <b, b>$, and evaluate similarly for $||a - b||^2$.

Proof. (Correct proof of the above lemma)
To show that $(y_n)_n$ is a Cauchy sequence, we have to use the parallelogram identity

$$||x - \frac{y_n + y_m}{2}||^2 + ||\frac{y_n - y_m}{2}||^2 = \frac{1}{2}(||x - y_n||^2 + ||x - y_m||^2)$$

as both terms in RHS tend to $d(x, V)^2$, but the first term in LHS is at least $d(x, V)^2$. $\square$

Remark. This holds more generally if V is a closed convex subset of H.

Generalize (Hohn-Banach Lemma?) If V is a closed convex set and $x \notin V$, then there exists a linear form l on H bounde and $\exists c \in \mathbb{R}$ s.t. $l(x) > c$ and $l(y) < c$ for all $y \in V$.

Corollary. Every closed subspace has an orthogonal complement, namely if $V \subseteq H$ is a closed vector subspace, then $H = V \oplus V^{\perp}$, where we write $V^{\perp} = \{x \in H, <x, y> = 0 \forall y \in V\}$. Note that $V^{\perp}$ is a closed subspace: $x_n \to x$, $<x_n, y> = 0 \implies <x, y> = 0$.

Proof. $V \cap V^{\perp} = \{0\}$ because $<x, x> = 0 \implies x = 0$. $H = V + V^{\perp}$ because $x - y \in V^{\perp}$ if y is the projection of x to V given by the lemma: this is because for all $z \in V$, $||x - y - z||^2 \geq ||x - y||^2$ since $z + y \in V$, i.e. $<x-y-z, x-y-z> \geq ||x-y||^2$, but LHS $= ||x-y||^2 + ||z||^2 + 2Re <x-y, z>$. So $\forall z \in V$, $||z||^2 + 2Re <x-y, z> \geq 0$. $\forall t > 0$, $t^2||z||^2 + 2tRe <x-y, z> \geq 0$, $t||z||^2 + 2Re <x-y, z> \geq 0$. When $t \to 0$, $Re <X-y, z> \geq 0 \forall z \in V$. Now $z \to -z$ gives $Re <x-y, z> = 0$, $z \to oz$ gives $im <X-y, z> = 0$, so $<x-y, z> = 0$, i.e. $x - y \in V^{\perp}$. □

Let E be a Bamach space (=complete normed vector space), $E^* =$dual of $E := \{$ bounded functionals on $E\} = \{l : E \to \mathbb{C}$ linear, $\sup_{||x|| \leq 1} |l(x)| < \infty\}$. Then E^* is also a Banach space (completeness follows that of $\mathbb{C}$), with norm $||l|| = \sup_{||x|| \leq 1} |l(x)|$.

Proposition. (Riesz representation theorem) If $E = \mathcal{H}$ is a Hilbert space, then $\mathcal{H}$ is self-dual, i.e. $\mathcal{H} \cong \mathcal{H}^*$, namely the map $\mathcal{H} \to \mathcal{H}^*$ by $y \to (l_y : x \to <x, y>)$ is an isomorphism, i.e. $\forall l \in \mathcal{H}^*, \exists y \in \mathcal{H}$ s.t. $l(x) = <x, y>$.

Proof. Let $V = \ker l$, V is closed because l is bounded. So by previous corollary $\mathcal{H} = V \oplus V^{\perp} = \ker l \oplus (\ker l)^{\perp}$. Now pick $x_0 \in (\ker l)^{\perp}$ s.t. $l(x_0) = 1$. Clearly x_0 is unique because if $l(x_1) = 1$ then $l(x_0 - x_1) = 0$, $x_0 - x_1 \in (\ker l)^{\perp} \cap \ker l = 0$. Now $l(y) - <y, \frac{x_0}{||x_0||^2}>$ must vanish on $\ker l$ and $(\ker l)^{\perp}$, so is zero everywhere. So $l(y) = <y, \frac{x_0}{||x_0||^2}>$. □

Proposition. (Jensen's inequality) Let $\Omega \subseteq \mathbb{R}^d$ be a convex set. Let $\phi : \Omega \to \mathbb{R}$ be a convex (continuous) function. Let X be an integrable $\mathbb{R}$-valued random variable s.t. $X \in \Omega$ a.s.. Then $\mathbb{E}(\phi(x)) \geq \phi(\mathbb{E}, X)$.

Remark. Ω convex means: for all $x, y \in \Omega$, $tx + (1-t)y \in \Omega \ \forall t \in [0, 1]$; ϕ convex means $\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y) \ \forall t \in [0, 1]$.

Proof. (sketch)
We first show $\mathbb{E}X \in \Omega$: use the geometric Hahn-Banach separation theorem: if $x \in \Omega$, then there exists l affine linear form $(l(x) = \sum a_i x_i + c, l(x) > 0, l(z) < 0 \ \forall z \notin \Omega$. If $\mathbb{E}X \notin \Omega$ then $l(\mathbb{E}X) < 0$, but $\mathbb{E}l(X) \geq 0$, impossible.

Then let $\mathcal{F}$ be the set of affine linear forms l s.t. $\phi(x) \geq l(x)$ for all $x \in \Omega$. Then we claim that $\phi(x) = \sup_{l \in \mathcal{F}} l(x)$ for all $x \in \Omega$ (ϕ continuous). To prove that we use the geometric Hahn-Banach separation theorem and apply it to $\Omega_{\phi} = \{(x, y), y \geq \phi(x)\}$ which is a convex set.

Proof of Jensen's inequality: pick $l \in \mathcal{F}$ s.t. $l(\mathbb{E}X) = \phi(\mathbb{E}X)$, which is also equal to $\mathbb{E}(l(x)) \leq E\phi(X)$ because $l \in \mathcal{F}$. □

Remark. If ϕ is convex on Ω then ϕ is continuous at every point in the interior (exercise).

Using this remark and an induction on d, one can remove the continuity assumption on ϕ.

Example. • $\phi = -\log x$ on $(0, +\infty)$ is convex. So if X is a real valued r.v., $X > 0$ a.s., then $\mathbb{E}(\log X) \leq \log(\mathbb{E}X)$.

• $x \to x^\alpha$, $\alpha \geq 1$ is convex. $\mathbb{E}(|X|^\alpha) \geq \mathbb{E}(|X|)^\alpha$. In particular, if $q \geq p$, then $\mathbb{E}(|X|^p)^{1/p} \leq \mathbb{E}(|X|^q)^{1/q}$. Hence $L^q(\Omega, \mathcal{A}, \mathbb{P}) \subseteq L^p(\Omega, \mathcal{A}, \mathbb{P})$.

Careful: in infinite measure e.g $\Omega = \mathbb{R}$, $\mu =$Lebesgue, then there is no inclusion relation between L^p and L^q.

2.3 Conditional Expectation

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra.

Proposition. $\forall A \in \mathcal{F}$, $\exists$ a unique $\mathcal{G}$-measurable R.V. denoted by $\mathbb{P}(A \mid \mathcal{G})$ s.t.

$$\forall B \in \mathcal{G} \; \mathbb{P}(B \cap A) = \mathbb{E}(1_B \mathbb{P}(A \mid \mathcal{G}))$$

and this is called the conditional probability of A w.r.t. $\mathcal{G}$.

We also say $\mathbb{P}(A \mid \mathcal{G})$ is the "probability of A knowing $\mathcal{G}$".

Example. Let $\mathcal{G} = \{B, B^c, \phi, \Omega\}$. Then

$$\mathbb{P}(A \mid \mathcal{G})_{(\omega)} = \left\{ \begin{array}{ll} \mathbb{P}(A \cap B)/\mathbb{P}(B) & \omega \in B \\ \mathbb{P}(A \cap B^c)/\mathbb{P}(B^c) & \Omega \in B^c \end{array} \right.$$

Notation: $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

If A is independent of $\mathcal{G}$, then $\mathbb{P}(A \mid \mathcal{G})(\omega) = \mathbb{P}(A)$.

Proof. $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Also they are Hilbert spaces, so $\exists Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ s.t. $1_A - Y \perp L^2(\Omega, \mathcal{G}, \mathbb{P})$, because $L^2(\mathcal{F}) = L^2(\mathcal{G}) \oplus L^2(\mathcal{G})^\perp$. But this means $\mathbb{E}(1_B(1_A - Y)) = 0 \; \forall B \in \mathcal{G}$, i.e. $\mathbb{P}(B \cap A) = \mathbb{E}(1_A Y)$ so we showed existence.

For uniqueness, suppose Y, Y' ($\mathcal{G}$-measures) satisfy $\mathbb{E}(1_B Y) = \mathbb{E}(1_B Y') \; \forall is \in \mathcal{G}$. But then $\mathbb{E}(1_B(Y - Y')) = 0$, so $Y = Y'$ a.s.. $\qquad \square$

More generally,

Proposition. Let X be an integrable r.v. on $L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}$ a sub-σ-algebra. Then there exists a unique $\mathcal{G}$-measurable r.v. Y s.t. $\mathbb{E}(1_B X) = \mathbb{E}(1_B Y) \; \forall B \in \mathcal{G}$.

Remark. Y is denoted by $\mathbb{E}(X \mid \mathcal{G})$ and is called the conditional expectation of X w.r.t. $\mathcal{G}$. If $X = 1_A$, then $Y = \mathbb{P}(A \mid \mathcal{G})$.

2.4 Convolutions

Definition. If f, g are functions in $L^1(\mathbb{R}^d)$, we define the convolution of f and g by

$$f * g(x) = \int_{\mathbb{R}^d} f(x-y)g(y)dy$$

whee dy is a Lebesgue measure.

Remark. This shows that $f * g \in L^1(\mathbb{R}^d)$ and $||f * g||_1 \le ||f||_1 ||g||_1$.

We say $L^1(\mathbb{R}^d)$ is a Banach algebra: Banach space with a product satisfying the above inequality.

Note it is commutative: $f * g = g * f$.

Definition. More generally, one can define the convolution of two measures on $\mathbb{R}^d$.

If μ, ν are probability measures, then we define $\mu * \nu$ to be the image of $\mu \otimes \nu$ on $\mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R}^d$ under the addition map $a : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ by $(x, y) \to x + y$. $\mu * \nu = a_*(\mu \otimes \nu)$, $A \subseteq \mathbb{R}^d$, $\mu * \nu(A) = \mu \otimes \nu * a^{-1}(A)) = \mu \otimes \nu\{(x, y) : x + y \in A\} = \int 1_A(x + y)d\mu(x)d\nu(y)$. We have $\mu * \nu(A) = \int \mu(A - y)d\nu(y) = \int \nu(A - x)d\mu(x)$.

Remark. This is consistent with the previous definition: $d\mu = fdx$, $f\nu = gdx$, $f, t \in L^1(\mathbb{R}^d)$, then $d(\mu * \nu) = f * gdx$.

If X, Y are independent $\mathbb{R}^d$-valued random variables with law μ_x and μ_y respectively, then $X + Y$ has law $\mu_x * \mu_y$.

Lemma. (continuity of translation on L^p)
If $p \in [1, \infty)$, and if $f \in L^p(\mathbb{R}^d)$, then $||\tau_h(f) - f||_p \to 0$ as $h \to 0$, where $\tau_h(f)(x) := f(x + h)$.

Proof. Step 1: Assume $f \in C_c(\mathbb{R}^d)$. Then it's clear by dominated convergence because $\tau_h(f) \to f$ pointwise as $h \to 0$.

Step 2: General case. Approximate f by a function $g \in C_c(\mathbb{R}^d)$, i.e. $\forall \varepsilon \exists$ such a g s.t. $||f - g||_p < \varepsilon$. Then $||\tau_h f - f||_p \le ||\tau_j f - \tau_h g||_p + ||\tau_h g - g||_p + ||g - f||_p \le 2\varepsilon + ||\tau_h g - g||_p$. $\qquad \square$

Lemma. (gaussian approximation)
If $p \in [1, \infty)$, and if $f \in L^p(\mathbb{R}^d)$, then $||f * g_\tau - f||_p \xrightarrow{\sigma \to 0} 0$, where g_σ is the density of a gaussian $N(0, \sigma^2)$,

$$g_\sigma(x) = (\frac{1}{\sqrt{2\pi\sigma^2}})^d e^{-\frac{x_1^2 + \ldots + x_d^2}{2\sigma^2}}$$

and $g_\sigma(x)dx =$ law of $\sigma(X_1, ..., X_d)$, where X_i is IID $N(0, 1)$.

Remark. $f * g_r$ is C^∞ (differentiation under integral sign).

Proof. $f * g_r(x) - f(x) = \mathbb{E}(f(x - \sigma\mathbf{X}) - f(x))$ where $\mathbf{X} = (X_1, ..., X_d)$, X_i's IID $N(0,1)$. By Jensen's inequality for $x \to x^p$,

$$||f * g_\sigma - f||_p^p \leq \mathbb{E}(||f(x - \sigma\mathbf{X}) - f(x)||_p^p)$$

then let $\sigma \to 0$ and use the previous lemma and Dominated Convergence. $\quad\square$

2.5 Fourier transform on $\mathbb{R}^d$

Definition. If $f \in L^1(\mathbb{R}^d)$, we can define its Fourier transform by

$$\hat{f}(u) = \int f(x) e^{i<u,x>} dx$$

where $<u, x> = u_1 x_1 + ... + u_d x_d$.

Remark. $|\hat{f}(u)| \leq ||f||_1$, so $\hat{f}$ is bounded;
$\hat{f}(u)$ is continuous in u (by Dominated Convergence).

More generally,

Definition. If μ is a probability measure on $\mathbb{R}^d$, we can define $\hat{\mu}(u) = \int e^{i<u,x>} d\mu(x)$.

Example. If X is on $\mathbb{R}^d$-valued r.v. with law μ_x, then $\hat{\mu}_X * u)$ is called the characteristic function of X.

Example. $\bullet$ $X \equiv c \in \mathbb{R}^d$ a.s. constant, then $\mu_x = \delta_u$ the dirac mass, and $\hat{\mu}_X(u) = e^{i<u,c>}$.
$\bullet$ $X =$ a normalized gaussian $N(0,1)$, $\mu_x = g(x) dx$ where $g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Then $\hat{\mu}_X(u) = e^{-u^2/2}$.

In particular, $\hat{\mu}_X(u) = \hat{g}(u) = \sqrt{2\pi} g(u)$, i.e. this is fixed point of the fourier transform.

Proof.

$$\hat{g}(u) = \int g(x) e^{iux} dx \int e^{-x^2/2} e^{iux} dx$$

$$\frac{d}{du} \hat{g}(u) = \int g(x) ix e^{iux} dx$$

but $-xg(x) = g'(x)$. Integrate by parts, the above is

$$= -\int g'(x) i e^{iux} = u \int g(x) e^{iux} dx = u\hat{g}(u)$$

$g'(u) = u\hat{g}$ so we solve this 1st order ODE and get $\hat{g}(u) = \hat{g}(0) e^{-u^2/2}$, so $\hat{g}(u) = e^{-u^2/2}$. $\quad\square$

Theorem. (Fourier Inversion Formula)
If f and $\hat{f}$ are in $L^1(\mathbb{R}^d)$, then $f(x) = \frac{1}{(2\pi)^d} \hat{\hat{f}}(-x)$ a.e., meaning f decomposes as a (continuous) linear combination of "characters", i.e. $\chi_u := x \to e^{i<u,x>}$. So

$$f(x) = \frac{1}{(2\pi)^d} \int \hat{f}(u) e^{-iux} du$$

Last time we introduced:

Theorem. (Fourier Inversion)
Suppose $f \in L^1(\mathbb{R}^d)$, and assume $\hat{f} \in L^1(\mathbb{R}^d)$. Then $f(x) = \frac{1}{(2\pi)^d}\hat{\hat{f}}(-x) = \frac{1}{(2\pi)^d}\int \hat{f}(u)e^{-i<u,x>}du$.
$<u,x> = u_1 x_1 + ... + u_d x_d$.

Remark. The functions $x \to e^{i<u,x>}$ are continuous group homomorphisms from $\mathbb{R}^d$ to $\{z \in \mathbb{C}, |z| = 1\}$.
Exercise: every continuous group homomorphism from $\mathbb{R}^d$ to U(the unit circle in complex plane) is of this form.

Remark. If $f, g \in L^1(\mathbb{R}^d)$ then $f * g = \int f(x-y)g(y)dy$ is s.t. $\widehat{f * g} = \hat{f} \cdot \hat{g}$.
To show this, we have

$$\int f * g(x)e^{i<u,x>}dx = \int\int f(x-y)g(y)e^{i<u,x>}dxdy$$

$$= \int\int f(z)g(y)e^{i<u,y+z>}dz$$

$$= \hat{f}(u)\hat{g}(u)$$

In probabilistic terms, we view $X \sim \mu_X$, $y \sim \mu_Y$, then $\hat{\mu}_X(u) = \mathbb{E}(e^{i<u,x>})$, $\hat{\mu}_Y(u) = E(e^{i<u,y>})$. X and Y are independent, so $\mathbb{E}(e^{i<u,X+Y>})$ is the product of the two expectations. Then rewrite the expectations as fourier transforms.

Proof. (of Fourier Inversion)
• First note that if it holds for $f \in C^1(\mathbb{R})$, it also holds for $f_t : x \to f(t_x)$ for $t > 0$ (i.e. dilation) by change of variables: $\hat{f_t}(u) = \int f(tx)e^{i<u,x>}dx = \frac{1}{t^d}\int f(tx)e^{i<u/t,x>}d(tx) = \frac{1}{t^d}\hat{f}(\frac{u}{t})$. So $\int \hat{f_t}(u)e^{-i<u,x>}du = \frac{1}{t^d}\int \hat{f}(\frac{u}{t})e^{-i<u/t,x>}du = \int \hat{f}(U)e^{-i<u,tx>}du = (2\pi)^d f(tx) = (2\pi)^d f_t(x)$.
• The inversion formula holds for $f = G$ a normalized Gaussian:

$$G(x) = \frac{1}{(\sqrt{2\pi})^d}e^{-\frac{x_1^2 + ... + x_d^2}{2}} = g(x_1)...g(x_d)$$

where $g(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $\hat{g}(u) = e^{-u^2/2}$ by last time. This is because

$$\hat{G}(u) = \int G(x)e^{i<u,x>}dx = \prod_1^d \int g(x_i)e^{-iu_i x_i}dx_i = \prod e^{-u_i^2}/2$$

$$= e^{-\frac{u_1^2 + ... + u_d^2}{2}} = (2\pi)^{d/2}G(u),$$

$$\hat{\hat{G}}(x) = \int \hat{G}(u)e^{i<u,x>}du = (2\pi)^{d/2}\int G(u)e^{i<u,x>}du = (2\pi)^{d/2}\hat{G}(x)$$

$$= (2\pi)^d G(x) = (2\pi)^d G(-x)$$

by the previous steps, the formula holds for $G_\sigma \sim N(0, \sigma^2 Id) \sim \sigma(X_1, ..., X_d)$ where $G_\sigma(x) = \frac{e^{-\frac{x_1^2 + ... + x_d^2}{2\sigma^2}}}{(\sqrt{2\pi\sigma^2})^d)} = \frac{1}{\sigma^d}G(x/\sigma)$. Now $\hat{G}_\sigma(u) = \hat{G}(\sigma u) = e^{-\sigma\frac{(u_1^2 + ... + u_d^2)}{2}}$.

Note that $\hat{G}_\sigma(u) \to 1$ as $\sigma \to 0$. Also $f * G_\sigma \xrightarrow{\sigma \to 0} f$, so $\hat{f}\hat{G}_\sigma \to \hat{f}$.

• Special case: f has the form $f = \phi * G_\sigma$ for some $\sigma > 0$ for some $\phi \in C^1(\mathbb{R}^d)$. Then $\hat{f} = \hat{\phi} \cdot \hat{G}_\sigma$. So

$$
\begin{aligned}
f(x) = \phi * G_\sigma(x) &= \int \phi(x - y)G_\sigma(y)dy \\
&= \int \phi(x - y) \int \hat{G}_\sigma(u)e^{-i<u,y>}\frac{du}{(2\pi)^d}dy \\
&= \int \int \phi(z)\hat{G}_\sigma(U)e^{-i<u,x-z>}\frac{dudz}{(2\pi)^d} \\
&= \int \underbrace{\hat{\phi}(u)\hat{G}_\sigma(u)}_{\phi * \hat{G}_\sigma = \hat{f}} e^{-i<u,x>}\frac{du}{(2\pi)^d}
\end{aligned}
$$

• For the general case, when $f, \hat{f}$ are in L^1, let $f_\sigma = f * G_\sigma$. We apply the Fourier Inversion formula to f_σ. So

$$
f_\sigma(x) = \int \hat{f}_\sigma(u)e^{-i<u,x>}\frac{du}{(2\pi)^d} = \int \hat{f}(u)\hat{G}_\sigma(u)e^{-i<u,x>}\frac{du}{(2\pi)^d}
$$

However, the last term tends to $\int \hat{f}(u)e^{-i<u,x>}\frac{du}{(2\pi)^d}$ pointwise a.e. x as $\sigma \to 0$ since $\bar{f} \in C^1$ by Dominated Convergence. Also, $f_\sigma(x)$ tends to f as $\sigma \to 0$ by the gaussian approximation lemma. So we have, a.e.,

$$
f(x) = \int \hat{f}(u)e^{-i<u,x>}\frac{du}{(2\pi)^d}
$$

$\square$

Theorem. (a) If $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, then $\hat{f} \in L^2(\mathbb{R}^d)$, and $||\hat{f}||_2 = (2\pi)^{d/2}||f||_2$ (Plancherel formula).
(b) Furthermore, if $f, g \in L^1 \cap L^2(\mathbb{R}^d)$, then $< \hat{f}, \hat{g} >= (2\pi)^d < f, g >$, where $< f, g >= \int_{\mathbb{R}^d} f\bar{g}dx$.
(c) Finally, the map $\mathcal{F} : L^1 \cap L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ by $f \to \frac{1}{(2\pi)^{d/2}}\hat{f}$ extends uniquely to a linear involutive isometry of $L^2(\mathbb{R}^d)$. Here, extends means to $L^2 \to L^2$, linear means $\mathcal{F}(f + \lambda g) = \mathcal{F}f + \lambda\mathcal{F}g$, involutive means $\mathcal{F} \circ \mathcal{F} = id_{L^2}$, and isometry means $||\mathcal{F}f||_2 = ||f||_2$, or $< \mathcal{F}f, \mathcal{F}g >=< f, g >$.

Proof. Proof of (a) and (b): First assume $f, \hat{f} \in L^1$. Then

$$
||\hat{f}||_2^2 := \int \hat{f}(u)\overline{\hat{f}(u)}du = \int [\int f(x)e^{i<x,u>}dx]\overline{\hat{f}(u)}du
$$

but $(x, u) \to f(x) \hat{\ } \overline{(Ĳf)}$ is in $L^1(\mathbb{R}^d \times \mathbb{R}^d)$. So Fubini applies, and above is equal to $\int \int f(x)(\int \hat{f}(u)e^{-i<u,x>}du)dx$. Apply Fourier inversion formula, this is equal to $\int f(x)f(\bar{x})(2\pi)^ddx = (2\pi)^d||f||_2^2$, and this is finite. So $\hat{f} \in L^2(\mathbb{R}^d)$ and

$||\hat{f}||_2^2 = (2\pi)^d ||f||_2^2$.

Similarly, if $f, g \in L^1 \cap L^2$, s.t. $\hat{f}, \hat{g} \in C^1$, the same calculation will show that $< \hat{f}, \hat{g} >= (2\pi)^d < f, g >$.

Now assume only $f \in L^1 \cap L^2$,, and no assumption on $\hat{f}$ (general case). Then consider gaussian approximation $f_\sigma := f * G_\sigma$. Now $\hat{f}_\sigma = \hat{f}\hat{G}_\sigma \in L^1$. So (a) holds for f_σ, and $f_\sigma \xrightarrow{\sigma \to 0} f$ in L^2 by Gaussian approximation lemma, $\hat{f}_\sigma = \hat{f}\hat{G}_\sigma$, $\hat{G}_\sigma$ tends to 1 pointwise as $\sigma \to 0$ ($||\hat{f}_\sigma||_2^2 = (2\pi)^d ||f_\sigma||_2^2 ||\hat{f}||_2^2 = (2\pi)^f ||f||_2^2$).

The proof to this is not that important, instead the formula itself is much more important.

(b) is basically the same thing.

(c) by (a) and (b), $||\mathcal{F}f||_2 = ||f||_2$ for all $f \in L^1 \cap L^2(\mathbb{R}^d)$. But $L_1 \cap L_2(\mathbb{R}^d)$ is dense in $L_2(\mathbb{R}^d)$, and the intersection contains $C_c(\mathbb{R}^d)$. $\square$

Define for $f \in L^2(\mathbb{R}^d)$, $\mathcal{F}f = \lim_{n \to \infty} \mathcal{F}f_n$ in $L^2(\mathbb{R}^d)$ for some (or any) $f_n \in L^1 \cap L^2$, $f_n \xrightarrow{L^2} f$ as $n \to \infty$.