MEASURE-THEORETIC PROBABILITY I

ABSTRACT. The goal of this course is to get answer the question: How to measure a chance?

CONTENTS

1. 1/4: Progression toward Jordan Measure

To get to know what is a measure, we start as history does. That is, we start with Jordan measure.

Def 1.1. *A subset of $\mathbb{R}_n$ is called elementary if it is a finite union of boxes, i.e. $B = I_1 \times I_2 \times \cdots \times I_n$ for any kind of intervals of $\mathbb{R}$.*

Also, given such a box B we can define its volumn by

$$|B| = \prod_{i=1}^{n} |a_i - b_i|$$

where

$$(a_i, b_i) \subset I_i \subset [a_i, b_i].$$

Def 1.2. *A Boolean Algebra on a set X is a family $\mathcal{B}$ of subsets of X which has*

(1) $\emptyset \in \mathcal{B}$;
(2) $\mathcal{B}$ is stable under finite union and complement.

Note that the above assumptions implies that if $A, B \in \mathcal{B}$, then $A \cap B \in \mathcal{B}$, $A \backslash B \in \mathcal{B}$, and $A \Delta B = (A \cup B) \backslash (A \cap B) \in \mathcal{B}$.

Example 1.3.

(1) Trivial algebra: $\{\emptyset, X\}$;
(2) Discrete Boolean algebra: $\mathcal{B} = 2^X$;
(3) For X a topology space, $\mathcal{B} =$ finite union of locally closed sets.

Def 1.4. *A finitely additive measure $(X, \mathcal{B})$ is the function $\mu : \mathcal{B} \to [0, \infty]$ such that*

(1) $\mu(\emptyset) = 0$*;*
(2) $\mu(E \uplus F) = \mu(E) + \mu(F)$ *where* $\uplus$ *is the disjoint union.*

Corollary 1.5. *For $E, F \in \mathcal{B}$, we have*

(1) (Sub-additivity): $\mu(E \cup F) \leq \mu(E) + \mu(F)$*;*
(2) (Monotonicity): $\mu(E) \leq \mu(F)$ *if* $E \subset F$.

Example 1.6.

(1) Counting measure: on $\mathcal{B} = 2^X$, we can let $\mu(E) := \#E$;
(2) More generally, for $f : x \mapsto [0, \infty]$, we can let $\mu(E) := \sum_{e \in E} f(e)$.

Proposition 1.7. *Let B be a box as defined above, and denote by $\mathcal{E}(B)$ the family of all elementary subsets of B, then*

(a) $\mathcal{E}(B)$ *is a Boolean Algebra.*
(b) $E \in \mathcal{E}(B)$ *is a finite union of disjoint boxes.*
(c) if $E \in \mathcal{E}(B)$ *is written in 2 different ways (2 different representations), i.e.*

$$E = \bigcup_{i=1}^{N} B_i = \bigcup_{j=1}^{k} B'_j$$

then

$$\sum_{i=1}^{N} |B_i| = \sum_{j=1}^{k} |B'_j|.$$

Proof. (a): Check that $\mathcal{E}(B)$ is a Boolean Algebra. We only do when dimension is 1, but really nothing has changed for higher dimensions.

It is stable under finite unions because the union of 2 elementary sets are just unions of two finite collections of boxes; it is stable under complement since we can just cut the interval into parts and choose the other ones.

(b): We do it via induction on the number of boxes. (note that here it really is ok for E to have different representations, as they will all be finite by definition).

If the representation of E is $E = B$, then we are done.

For the induction step, we have

$$E = \bigcup_{i=1}^{N-1} B_i \cup B_N = \biguplus_{j=1}^{k} B'_j \cup B_N = \biguplus_{j=1}^{k} (B'_j \backslash B_N) \uplus B_N.$$

(c): under the condition of 2 representations, using a subset argument we can get

$$E = \biguplus_{i,j} (B_i \cap B'_j)$$

and it is enough to consider when $N = 1$, i.e. $E = B$ is a box. We do it only in 2d. The way to do it is to refine the partition into smaller boxes whose sides are pieces of I'_j, and with some tedious refinement of grids we will be done. (skip details). □

Def 1.8. *We may set* $\mu(E) = \sum_{i=1}^{N} |B_i|$ *where* B_i *is a representation of* E*.*

Proposition 1.9. *As defined above,* μ *is a finitely additive measure on* $(B, \mathcal{E}(B))$*.*

(prove next time.q)

Def 1.10. *A subset* A *of* $\mathbb{R}^n$ *is called* <u>*Jordan Measurable*</u> *if* $\forall \varepsilon > 0$, $\exists$ *elementary subsets* E, F *such that* $E \subset A \subset F$ *with*

$$\mu(E \backslash F) < \varepsilon.$$

If A is Jordan measurable, we define its <u>*Jordan measure*</u> *by*

$$\mu_J(A) := \inf_{\substack{A \subset F \\ F-elementary}} \mu(F).$$

Theorem 1.11. *If we fix a box in* $\mathbb{R}^n$*, then the collection of Jordan measurable sets* $J(B)$ *inside* B *is a Boolean algebra and* μ_J *is a finitely additive measure on it.*

Proof. $\mathcal{J}(B)$ is stable under finite unions by taking unions of corresponding E and F. Moreover, it's stable under complement since $\mu(E^c \backslash F^c) < \varepsilon$... The rest are easy exercise. □

2. 1/9: LEBESGUE MEASURE

Before we went into Lebesgue measure, let's have a few remarks.

Remark 2.1.

An equivalent definition of A being Jordan Measurable is that $\forall \varepsilon > 0$, there exists a finite union of boxes $F = \bigcup_{i=1}^{N} B_i$ and $F \backslash A$ is contained in an elementary set of measure less than ε.

Proposition 2.2. *Let A and A' be disjoint sets in $\mathcal{J}(B)$, then*

$$\mu(A \uplus A') = \mu(A) + \mu(A').$$

Proof. We prove in the two directions. But first we fix an $\varepsilon > 0$.

For the less than direction, we know by remark 2.1 that there exists elementary sets F, F' such that $A \subset F$, $A' \subset F'$ such that

$$\begin{cases} \mu(F) \leq \mu(A) + \varepsilon \\ \mu(F') \leq \mu(A') + \varepsilon \end{cases}$$

This implies that

$$\mu(A \uplus A') \leq \mu(F \cup F') \leq \mu(A) + \mu(A') + 2\varepsilon$$

where we used the subadditive property, so no need of F, F' being disjoint.

Now for the larger than direction, we can find (using the other direction of the remark) disjoint sets E, E' such that $E \subset A$ and $E' \subset A'$ for which we have

$$\begin{cases} \mu(e) \geq \mu(A) + \varepsilon \\ \mu(E') \geq \mu(A') + \varepsilon \end{cases}$$

This implies that

$$\mu(A \uplus A') \geq \mu(E \uplus E') = \mu(E) + \mu(E') \geq \mu(A) + \mu(A') + 2\varepsilon$$

where we really need E, E' to be disjoint. Yet this is very obvious since they are contained in A and A'.

But ε is arbitrary so

$$\mu(A \uplus A') = \mu(A) + \mu(A').$$

□

Now why do we bother with Lebesgue measure at all if we have Jordan measure? The answer is that

(1) Any unbounded sets are not Jordan Measurable.

(2) Some "simple minded" bounded sets can be not Jordan measurable, for instance, let $A := \mathbb{Q} \cap [0,1]$, then it's not Jordan measurable.

To see this, assume the contrary, then for any ε $\exists$ elementary sets $E \subset A \subset F$ such that $\mu(F - E) < \varepsilon$.

But E has to be a finite collection of points since it is contained in A, and F must contain $[0,1]$ with possibly countable points left out since $\mathbb{R} - F$ must be form of intervals. So contradiction as $\mu(F) = 1, \mu(E) = 0$.

(3) Lebesgue measure just behaves super well under taking limits. This is also a very essential reason, though not as explicit as the first 2 reasons.

Now we move into the lecture of Lebesgue measure.

Def 2.3. *$\forall \mathbb{E} \in \mathbb{R}^d$, we can define its Lebesgue <u>outer measure</u> by*

$$\mu^*(E) := \inf \left\{ \sum_{n \geq 1} |B_n| \middle| E \subset \bigcup B_n \right\}$$

where B_n are boxes.

Def 2.4. *A subset $E \subset \mathbb{R}^d$ is <u>Lebesgue measurable</u> if $\forall \varepsilon > 0$, $\exists C := \bigcup_{n \geq 1} B_n$ such that $\mu^*(C \setminus E) < \varepsilon$. We denote this collection by $\mathcal{L}$.*

Remark 2.5.

(1) We can always assume boxes are open here.
(2) Jordan measurable $\Rightarrow$ Lebesgue measurable.

Proposition 2.6.

(a) μ^* extends μ, namely $\mu^* = \mu(E)$ for $E \in \mathcal{J}(B)$.
(b) The family $\mathcal{L}$ forms a Boolean algebra that is stable under countable unions.
(c) μ^* is countably additive on $(\mathbb{R}^d, \mathcal{L})$, i.e.

$$\mu^*\left(\bigcupdot_{n \geq 1} E_n \in \mathcal{L} \right) = \sum_{n \geq 1} \mu^*(E_n).$$

As for the proof, (a) and (b) are homework questions. And (c) is proven later in next class.

Lemma 2.7. *The function μ^*, by definition, is*

(a) monotone: if $A \subset B$, then $\mu^*(A) \leq \mu^*(B)$.
(b) countably subadditive, that is, for any countable family $A_n \in \mathcal{L}$ we have

$$\mu^*\left(\bigcup_{n \geq 1} A_n \right) \leq \sum_{n \geq 1} \mu^*(A_n)$$

Proof. (a) is obvious by definition so we only prove (b) here.

(b):

We pick $\varepsilon > 0$, $C_n := \bigcup_{i\geq 1} B_{n,1}$ such that $A_n \subset C_n$ and

$$\sum_{i\geq 1} |B_{n,i}| \leq \mu^*(A_n) + \frac{\varepsilon}{2^n}$$

we can do this since $A_n \in \mathcal{L}$. Now we can get

$$\mu^*\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n,i\geq 1} |B_{n,i}| \leq \mu^*(A_n) + \varepsilon \cdot \left(1 + \frac{1}{2} + \frac{1}{4} + \dots\right)$$

where since ε is arbitrary we get what we want. $\square$

Lemma 2.8. *Let $A = \bigcap_{n\geq 1} E_n$ be a countable intersection of elementary sets such that $E_{n+1} \subset E_n$, then A is in $\mathcal{L}$ and*

$$\lim_{N\to\infty} \mu(E_N) = \mu^*(A).$$

Proof. For any n we have that $A \subset E_n$ and

$$E_n \backslash A = \bigcup_{i\geq n} E_i \backslash E_{i+1}.$$

Then sub additivity shows that

$$\mu^*(E_n\backslash A) \leq \sum_{i\geq 1} \mu^*(E_i\backslash E_{i+1}) = \sum_{i\geq n} \mu(E_i\backslash E_{i+1}) \leq \mu(E_n) < \infty$$

SO for fixed ε, $\exists N_0$ such that for all $N \geq N_0$

$$\sum_{i\geq N} \mu^*(E_k\backslash E_{k+1}) < \varepsilon$$

since the series $\sum_{i\geq n} \mu(E_i\backslash E_{i+1})$ is bounded by a finite number, and the sum above is just the tail of the series. Thus, $\mu^*(E_N\backslash A) < \varepsilon$ so that A is Lebesgue measurable.

Now, by subadditivity

$$\mu^*(A) \leq \mu^*(E_n) \leq \mu^*(E_n\backslash A) + \mu^*(A)$$

where we note that for $n \geq N$ the right part is just $\varepsilon + \mu^*(A)$. But this is just a squeeze theorem so if we add a limit to the inequality we get

$$\mu^*(A) \leq \lim_{n\to\infty} \mu^*(E_n) \leq \varepsilon + \mu^*(A)$$

which means what we want:

$$\mu^*(A) = \lim_{n\to\infty} \mu^*(E).$$

$\square$

This lemma is actually not in class, but it explains the proof of the corollary below.

Lemma 2.9. *Countable union of measurable sets is measurable.*

Proof. For each measurable set and fixed $\varepsilon > 0$, we can first label the sets and then find corresponding countable union of elementary sets such that the difference is less than $\frac{\varepsilon}{2^{i+1}}$. In other words, for

$$A := \bigcup_{n\geq 1} E_n$$

we can find $C_n := \bigcup_{i\geq 1} B_{n,i}$ for elementary sets $B_{n,i}$ such that $E_n \subset C_n$ and

$$\mu^*(C_n \backslash E_n) < \frac{\varepsilon}{2^{i+1}}.$$

Now, we can just using subadditivity to get

$$\mu^*\left(\left[\bigcup_{n\geq 1} C_n\right] \backslash A\right) = \mu^*\left(\bigcap_{k\geq 1}\left[\left(\bigcup_{k\geq 1} C_n\right) \backslash E_k\right]\right) \leq \sum_{n\geq 1} \mu^*(C_n \backslash E_n) \leq \sum_{n\geq 1} \frac{\varepsilon}{2^{i+1}}$$

where the first inequality is due to the fact that if a point x is in the set in the second expression, it is definitely in at least on of the sets in the sum. But then since ε is arbitrary we are done. □

Corollary 2.10. *Every open and every closed subsets of $\mathbb{R}^n$ is in $\mathcal{L}$.*

Proof. Every open set can be expressed as a countable union of open boxes. This can be done by first finding a countable box cover of any rational points in the set, then do it for all rational points. But this directly says that every open set is in $\mathcal{L}$.

If C is closed, then

$$C = \bigcup_{n} C \cap B_n$$

for $B_n = [-n, n]^d$. This means that C can be written as a countable union of sets. If all those sets are measurable, then we can only deal with the case where C is bounded.

But then $C \subset B$ for some open box B, which implies that $B \backslash C$ is open. By the first part of this theorem we can write it as countable union of open boxes

$$B \backslash C = \bigcup_{n\geq 1} \tilde{B}_n$$

and hence

$$C = \bigcap_{n\geq 1} (B \backslash \tilde{B}_n)$$

where it means that C is a countable intersection of elementary sets.

Now we use the lemma 2.4 above. We can construct a nesting sequence F_n from any collection of E_n using the map $F_n = E_1 \cap \cdots \cap E_n$ (each F_n is also elementary) where we get that $\bigcap_{n\geq 1} F_n = \bigcap_{n\geq 1} E_n$. Where since the left hand side is measurable by Lemma 2.4, the right hand side is. That is, every countable intersection of elementary sets is measurable, and so we are done. $\square$

Lemma 2.11. *Assume that $E \subset \mathbb{R}^d$ is Lebesgue measurable. Fix $\varepsilon > 0$, then $\exists$ closed subset F and open set U such that*

$$F \subset E \subset U \quad \textit{with} \quad \mu^*(U \backslash F) < \varepsilon.$$

Moreover, $\exists$ a set $N \in \mathcal{L}$ such that $E = B \backslash N$ and

$$\mu^*(N) = 0$$

for which we call the null set, and B is a countable intersection of open boxes.

Proof.

Let $U = \dot{\bigcup}_{i\geq 1} B_i$ be such that $\mu^*(U \backslash E) < \dfrac{\varepsilon}{2}$.

Yet for any $\varepsilon > 0$, let B_n' be open boxes that has $B_n \subset B_n'$ with $\mu(B_n' \backslash B_n) < \dfrac{\varepsilon}{2^{n+1}}$. This can be done because each box is bounded, and we are in $\mathbb{R}^n$, then we have

$$\mu\left(\left(\bigcup_{n\geq 1} B_n'\right) \Big\backslash \left(\dot{\bigcup_{n\geq 1}} B_n\right)\right) \leq \mu\left(\bigcup_{n\geq 1} (B_n' \backslash B_n)\right) < \sum_{n\geq 1} \frac{\varepsilon}{2^{n+1}} = \frac{\varepsilon}{2}$$

thus

$$\mu\left(\left(\bigcup_{n\geq 1} B_n'\right) \Big\backslash E\right) = \mu\left(\left(\bigcup_{n\geq 1} B_n'\right) \Big\backslash \left(\dot{\bigcup_{n\geq 1}} B_n\right)\right) + \mu^*(U \backslash E) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Now that we are done with the open set part, we use that result to prove the closed set part. Yet the key to it is nothing but to construct such a open larger set Ω with

$$\mu^*(\Omega \backslash E^c) < \frac{\varepsilon}{2}$$

and let $F = \Omega^c$. Then by disjoint addition

$$\mu^*(U \backslash F) = \varepsilon$$

hence we are done.

For the other part, we can find $U_n \supset E$ with $\mu^*(U_n \backslash E) \leq \dfrac{1}{n}$, then we claim that $B = \bigcap_{n\geq 1} U_n$ is the right B with $N = B \backslash N$ that satisfies the condition.

Indeed $\mu(B \backslash N) < \dfrac{1}{n}$ for all n and hence it is 0. $\square$

3. 1/12: More on Lebesgue Measure; General measure

We start by some leftover comments on Lebesgue measure. First, we finish what's left from last class: part (c) of proposition 2.2.

As a reminder, proposition 2.2 is a collection of basic properties of μ^*, the statements to prove is the following:

μ^* is countably additive on $(\mathbb{R}^d, \mathcal{L})$, i.e. for $E_n \in \mathcal{L}$

$$\mu^*\left(\bigsqcup_{n\geq 1} E_n\right) = \sum_{n\geq 1} \mu^*(E_n).$$

Proof. (Proposition 2.2 (c)):

We make the following simplifications:

(1) It's enough to prove the statement for bounded E_n;
(2) It's enough to prove that μ^* is finitely additive on bounded E_n;
(3) It's enough to show that $\mu^*(E \sqcup F) = \mu^*(E) + \mu^*(F)$ for E, F being countable intersections of elementary sets.

Proof of simplification (1):

For $B_n = [-n, n]^d$ with $X_n = B_n \backslash B_{n-1}$, then we have

$$E := \bigsqcup_{m,n\geq 1} E_n \cap X_m; \quad \mu^*(E_n) = \sum_{m\geq 1} \mu^*(E_n \cap X_m)$$

which means that

$$\mu^*(E) = \sum_{m,n\geq 1} \mu^*(E_n \cap X_m) = \sum_{n\geq 1} \mu^*(E_n)$$

and thus again by Lemma 2.5 we can be safe with only bounded sets.

Proof of simplification (2):

For E_n uniformly bounded, use the compactness of any closed box that contains the countable union to get a finite cover, on which we will get the same result.

Proof of simplification (3): This is straight forward.

To proof the statement required in simplification (3):

That the left is lesser than the right is by definition (one by infimum). We only prove that the right is smaller than the left. This can be done with the following idea: we can find a sequence of covers of E and F that for each box cover that covers $E \sqcup F$, it is contained in it up till a measure 0 set.

For each elementary set in the combination of E, we can find a finite union of boxes that represents it. For that finite union we can find a larger open box cover which is larger but with in ε volume for any ε. So we choose an absolutely converging series and use each term in the summand to scale the corresponding open box outer cover difference. In the end we can get $c\epsilon$ for any cover that covers E and F, respectively. So as ε is arbitrary, we are done.

This is very vague but it mimics the proof in the first homework.

□

At last, we define the Lebesgue measure:

Def 3.1. *μ^* on $\mathcal{L}$ is called the Lebesgue measure.*

Lemma 3.2. *(Translation invariant principle): For any vector v in the space and any $E \in \mathcal{L}$ we have*

$$\mu(E) = \mu(E + v).$$

The prove is nothing but to add v on each of E's outer covers.

So a summary of what we've learned about Lebesgue measure is the following:

(1) $\mathcal{L}$ is a Boolean Algebra that is stable under countable unions, i.e. a σ-algebra.
(2) Every open and every closed set is in $\mathcal{L}$.
(3) Every set can be approximated by

$$F \subset E \subset U$$

with F closed and U open such that $\mu(U \backslash F) < \varepsilon$.
(4) Translation invariant, as in lemma 3.1 above.

Before we go into general measure, we look at a set in $\mathbb{R}$ that's not Lebesgue measurable.

Example 3.3. *Vitali's counter example.*

Let $E \subset [0, 1]$ such that $\forall x, \exists! e \in E$ such that $x - e \in \mathbb{Q}$. With AC this is fine. We claim that such E is not Lebesgue measurable.

Proof.

First, fix $r_1, \dots, r_N \in \mathbb{Q} \cap [0, 1]$ and suppose that E is Lebesgue measurable. Then

$$\mu^* \left(\bigsqcup_{i \geq 1}^{N} (r_i + E) \right) = \sum_{i=1}^{N} \mu^*(r_i + E) = N \mu^*(E)$$

where the first union is disjoint due to the definition of E. Now since $r_i + E \subset [0, 2]$ for all i, we know that $\mu^*(E) = 0$, since otherwise we can find N large enough such that $N\mu^*(E)$ is larger than 2.

Yet another observation is that $\bigcup_{r \in \mathbb{Q}} (E + r) \supset [0, 1]$ since each $x \in [0, 1]$ is $q + e$ for some $e \in E$ and $q \in \mathbb{Q}$.

This means

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

$\square$

Actually $\mu^*(E)$, the outer measure of the example above is greater than 0, since any set with outer measure 0 is measurable.

3.1. **Abstract Measure Theory.**

Def 3.4. *A σ-algebra $\mathcal{F}$ on X is a Boolean algebra stable under countable unions.*

Def 3.5.

- *A measurable space is a pair $(X, \mathcal{A})$ where $\mathcal{A}$ is a σ-algebra on X.*
- *A measure on $(X, \mathcal{A})$ is a map $\mu : \mathcal{A} \to [0, +\infty]$ that satisfies*
  (1) $\mu(\emptyset) = 0$;
  (2) Countable additivity: $\mu\left(\bigsqcup E_n\right) = \sum \mu(E_n)$.
- *A measure space is the triple $(X, \mathcal{A}, \mu)$.*
- *A probability measure is the measure $\mathbb{P}$ such that $\mathbb{P}(X) = 1$ on $(X, \mathcal{A})$.*

Example 3.6.

(1) The Lebesgue measure $(\mathbb{R}^d, \mathcal{L}, \mu)$ is a measure space.
(2) For any $A_0 \in \mathcal{L}$, we can define a measure $\mu_0 : \mathcal{L} \to [0, +\infty]$ with

$$\mu_0(E) = \mu(A_0 \cap E).$$

Some easy properties of measure are:

Proposition 3.7. *Let $(X, \mathcal{A}, \mu)$ be a measure space, then*

(a) μ is monotone.
(b) μ is countably additive, i.e.

$$\mu\left(\bigcap_{n\ge1} E_n\right) \le \sum_{n\ge1} \mu(E_n).$$

(c) μ is upward monotone convergent, i.e. for $E_1 \subset E_2 \subset \dots$ that are in $\mathcal{A}$, we have

$$\mu\left(\bigcup_{n\ge1} E_n\right) = \lim_{n\to\infty} \mu(E_n) = \sup_n \mu(E_n).$$

(d) μ is downward monotone convergent, i.e. for $E_1 \supset E_2 \supset \dots$ that are in $\mathcal{A}$ ***and that*** *$\mu(\mathbf{E_1}) < \infty$, we have*

$$\mu\left(\bigcap_{n\ge1} E_n\right) = \lim_{n\to\infty} \mu(E_n) = \inf_n \mu(E_n).$$

Note that the downward monotone convergent requires that E_1 is bounded. A counter example that illustrates this requirement is the following: $E_n = [n, \infty)$. Then $\mu(E_n) = \infty$ yet $\mu\left(\lim_{n\to\infty} E_n\right) = 0$.

Proposition 3.8. *For any collection $\mathcal{F}$ of subsets, we can define the σ-algebra generated by $\mathcal{F}$ as the smallest σ-algebra containing $\mathcal{F}$. We denote it as $\sigma(\mathcal{F})$. In other words*

$$\sigma(\mathcal{F}) = \bigcap_{i\in\mathcal{I}}$$

such that σ_i is a σ-algebra with $\mathcal{F} \subset \sigma_i$.

Proof. First of all, 2^X is in the intersection. Then, we only need to show that the intersection of σ-algebras are still σ-algebra. But that's easy (actually in hw). $\square$

Def 3.9. *Let X also be a topological space, then the σ-algebra generated by all open sets of X is a Borel-algebra.*

Def 3.10. *A premeasure on a Boolean algebra $\mathcal{B}_0$ is a finite additive measure $\mu_0 : B \to$* $[0, \infty]$ *with*

- $\mu(\emptyset) = 0$;
- $\mu\left(\bigsqcup_{i=1}^k B_i\right) = \sum_{i=1}^k \mu_0(B_i)$ *for* $B_i \in \mathcal{B}_0$.

We end today's lecture with a higher level generation of what we did with Jordan measure and Lebesgue measure.

Theorem 3.11. *(Hahn Kolmogorov extension theorem): Every premeasure $\mu_0 : \mathcal{B}_0 \to$* $[0, \infty]$ *can be extended to a measure $\mu : \sigma(\mathcal{B}_0) \to [0, \infty]$.*

The proof of this mimics the construction of the Lebesgue measure that we did.

4. 1/18: More on general measures; Measurable functions

Theorem 4.1. *Let $\mathcal{P} \subset \Omega$ be a π-system. If μ_1 and μ_2 are measures on $\sigma(\mathcal{P})$, they agree on $\mathcal{P}$ and there exists a sequence of sets $A_1 \subset A_2 \subset \cdots \in \mathcal{P}$ such that $\cup A_i = \Omega$ and $\mu_{1,2}(A_i) < \infty$. Then μ_1, μ_2 agrees on $\sigma(\mathcal{P})$.*

Note that there is an addition condition on the theorem. We really need that since a counter example is provided in Tao's book. Check that.

To prove this theorem, we also need some preliminaries.

Def 4.2. *π-systems are system of sets that are closed under finite intersection.*

Def 4.3. *λ-systems are $\mathcal{L}$ that satisfies*

- $\Omega \in \mathcal{L}$;
- $A, B \in \mathcal{L}$, $B \backslash A \in \mathcal{L}$;
- *For* $A_1 \subset \cdots \in \mathcal{L}$, $\bigcup A_i \in \mathcal{L}$.

Theorem 4.4. *(Dynkin's $\pi - \lambda$ theorem) Let $\mathcal{P}$ be a π-system, $\mathcal{L}$ be a $\lambda-$system with $\mathcal{P} \subset \mathcal{L}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Def 4.5. *σ-finiteness: μ is σ-finite if $\exists$ a collection of sets $A_1 \subset A_2 \subset \ldots$ such that $\mu(A_i) < \infty$ and $\bigcup A_i = \Omega$.*

Proof. Of Theorem 4.1.

Step 1: Fix $A \in \mathcal{P}$, $\mu_1(A) = \mu_2(A) < \infty$. Let $\mathcal{L}$ be the collection of

$$\mathcal{L} := \{B \in \sigma(\mathcal{P}) : \mu_1(A \cap B) = \mu_2(A \cap B)\}$$

and we claim that $\mathcal{L}$ is a λ-system. Now by the $\pi - \lambda$ theorem we have that $\mathcal{L} = \sigma(\mathcal{P})$ since obviously $\mathcal{L} \subset \sigma(\mathcal{P})$. We have that proof of step 1 $\iff$ Hw2. prob 1, so we skip it here.

Now we take any $B \in \sigma(\mathcal{P})$ and we know for any $A \in \mathcal{P}$, $\mu_1(A) = \mu_2(A) < \infty$ and $\mu_1(A \cap B) = \mu_2(A \cap B)$ by above argument. Now by σ finiteness we pick $A_1 \subset A_2 \subset \ldots$ and get

$$\mu_1(B) = \lim_{n\to\infty} \mu_1(A_n \cap B) = \lim_{n\to\infty} \mu_2(A_n \cap B) = \mu_2(B).$$

$\square$

We restate the Hahn Kolmogorov extension theorem just because she likes to:

Theorem 4.6. *For Boolean algebra $\mathcal{B}_0$, every pre-measure $\mu_0 : \mathcal{B}_0 \to [0, \infty]$ can be extended to a countably additive measure $\mu : \mathcal{B} \to [0, \infty]$ where $\mathcal{B} = \sigma(\mathcal{B}_0)$. Moreover, the extension is unique if $A_1 \subset A_2 \subset \cdots \in \mathcal{B}_0$ such that $\Omega = \bigcup A_i$ and $\mu_0(A_i) < \infty$.*

Corollary 4.7. *The Lebesgue measure is the unique translation invariant measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ with $\mu([0, 1]^n) = 1$.*

The idea of the proof is the following: Let $\mathcal{F}$ be the family of all boxes in $\mathbb{R}^d$, so it's a π-system and $\sigma(\mathcal{F}) = \mathcal{B}(\mathbb{R}^n)$, the Borel algebra on $\mathbb{R}^d$.

We need to check that any 2 measures with the above properties agree on $\mathcal{F}$, which is in homework. A remark is that in this setting the σ-finiteness is satisfied.

Yet another remark is that $\sigma(\mathcal{J}) \subsetneq \mathcal{L}$, where $\mathcal{J}$ is Jordan measurable sets, $\mathcal{L}$ is the set of Lebesgue measurable sets.

4.1. **Measurable functions.**

Def 4.8. *Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A map $f : X \to Y$ is measurable if the preimage $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$. Where of course f^{-1} might not be a function.*

Proposition 4.9. *The composition of measurable functions is a measurable function. The identity map is a measurable function.*

Lemma 4.10. *Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable space with f a mapping between them. Now, suppose that $\mathcal{B} = \sigma(\mathcal{P})$, then if $f^{-1}(C) \in \mathcal{A}$ for $C \in \mathcal{P}$, then f is measurable.*

Proof. Consider the following set of $B \in \mathcal{B}$ such that $f^{-1}(B) \in \mathcal{A}$. We check that it is a σ-algebra. If it indeed is one, then the statement follows since $\sigma(\mathcal{P})$ is the smallest σ-algebra containing P, i.e. it is both smaller and larger than what we want, so it is what we want.

Now to show that it is a σ-algebra we just check property by property, which is routine. $\square$

Lemma 4.11. *Suppose X, Y are topological spaces, then any continuous function between $(X, \mathcal{B}(X))$ and $(Y, \mathcal{B}(Y))$ is measurable.*

Just by definition and the previous lemma we can get the proof.

Lemma 4.12. *Sums and products of real valued functions are measurable.*

Proof. We construct the following relations:

$$(X, \mathcal{A}) \to (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

with the following maps

$$x \mapsto (f(x), g(x)) \mapsto f(x) + g(x)$$

or

$$x \mapsto (f(x), g(x)) \mapsto f(x) \cdot g(x).$$

Now we just use the continuity to prove that the second map is measurable, for which we again use composition of measurable functions are measurable to get the result. $\square$

Proposition 4.13. *Suppose $f_n : (X, \mathcal{A}) \to \mathbb{R}^*$, the extended real line with $0 \cdot \infty = 0$, is a sequence of measurable functions, then $\inf_{n \geq 1} f_n(x)$ and $\sup_{n \geq 1} f_n(x)$ are measurable.*

Proof. Let $g(x) = \inf_{n \geq 1} f_n(x)$. Then for any $t \in \mathbb{R}^*$, we have

$$g^{-1}([t, \infty)) = \bigcap_{n=1}^{\infty} f_n^{-1}([t, \infty)) \in \mathcal{A}$$

since $g(x) \geq t$ implies $f_n(x) \geq t$.

Now we only need to check that $[t, \infty)$ generates the σ-algebra for $\mathbb{R}^*$, which is an exercise.

The same applies for *sup*. $\square$

Corollary 4.14. *For a sequence of measurable functions $\{f_n\} : (X, \mathcal{A}) \to \mathbb{R}^*$, $\limsup_{n\to\infty} f_n(x)$ and $\liminf_{n\to\infty} f_n(x)$ are measurable.*

The reason for the corollary is simply that lim sup is the composition of inf and sup.

Now since if the limit of f_n exists, then it is the lim sup, hence measurable.

Remark 4.15. *The image of a Borel set under a measurable function need not be measurable. It is not comforting, but the nature of things.*

Now we march a little bit towards Lebesgue integration. The first step is to define simple functions.

Def 4.16. *A non-negative simple function on $(X, \mathcal{A}, \mu)$ is a function of the form*

$$f := \sum_1^N a_i \mathbb{1}_{A_i}$$

with $a_i \in [0, \infty]$, $A_i \in \mathcal{A}$.

5. 1/23: Lebesgue integration; Convergence theorems

First we go and define the Lebesgue integration from the basics.

5.1. **Lebesgue Integration.**

Recall that an indicator function of a measurable set is a measurable function, so we have that simple functions are measurable functions.

An easy exercise is that simple functions does not depend on it's representation since it's just a finite sum and will remain finite and easy to deal with.

Def 5.1. *For f-simple, we define it's integral as*

$$\int_X f\,d\mu := \sum_{i=1}^{N} a_i \cdot \mu(A_i).$$

Now we can use the above definition to define the integral of non-negative measurable functions.

Def 5.2. *Let $f \geq 0$ be measurable, then*

$$\int_\Omega f\,d\mu = \sup_{g \in SF^+(f)} \int g(\omega)d\mu$$

where

$$SF^+(f) := \{s\text{-}simple | s(x) \leq f(x), \forall x\}.$$

Another exercise is that when f is a non-negative simple function, the definition 5.2 agrees with 5.1.

Now we further extend to all measurable functions.

Def 5.3. *For $f : \Omega \to \mathbb{R}^*$ measurable, define*

$$f^+ := \max\{f, 0\}, \quad f^- := -\min\{f, 0\}$$

which are both non-negative and measurable by a limsup argument. Then we say that the intergral of f is defined if at least one of the integrals $\int f^+ d\mu$ or $\int f^- d\mu$ are finite.

In this case, we define

$$\int f\,d\mu = \int f^+ d\mu - \int f^- d\mu.$$

Def 5.4. *The measurable function f is called* <u>*integrable*</u> *if the integrals $\int f^+ d\mu$ and $\int f^- d\mu$ are finite.*

Note that a function can have a integral even if it is not integrable.

Def 5.5. *For $S \in \mathcal{F}$, where $\mathcal{F}$ is the σ-algebra, we define*

$$\int_S f d\mu := \int_\Omega f \mathbb{1}_S d\mu$$

5.2. **Monotone convergence theorem.**

Theorem 5.6. *(Monotone Convergence theorem): Suppose $\{f_n\}_{n\geq 1}$ are measurable non-negative functions that increases point wise to a limit f, then f is measurable and*

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

We will prove the equality in two directions, as is expected. But one of the directions require a lot of work, for which we first prove the following lemma.

Lemma 5.7. *Let $S : \Omega \to [0, \infty]$ be a simple function (actually works for measurable, but not necessary here). For each $S \in \mathcal{F}$, define*

$$\nu(S) := \int_S s d\mu$$

then ν is a measure on $(\Omega, \mathcal{F})$.

Proof. (of Lemma 5.2)

Since s is simple, then we can write

$$s = \sum_{i=1}^n a_i \mathbb{1}_{A_i}.$$

To check the measure properties, we first note that by definition $\nu(\emptyset) = 0$.

To prove disjoint additive, let $S_1, S_2, \dots$ be a collection of disjoint sets and let $S := \sqcup S_i$. We can compute

$$\nu(S) = \sum_{i=1}^n a_i \mu(A_i \cap S) = \sum_{i=1}^n a_i \sum_{j=1}^\infty \mu(A_i \cap S_j) = \sum_{j=1}^\infty \sum_{i=1}^n a_i \mu(A_i \cap S_j) = \sum_{j=1}^\infty \nu(S_j)$$

where we can change the summation because everything is positive and thus well defined. □

Proof. (of Monotone convergence theorem)

Now there's 2 directions that we need to show, and we'll first do the easy one.

$\underline{\int f d\mu \geq \lim_{n\to\infty} \int_\Omega f_n d\mu:}$

This is just because $f \geq f_n$. Since by definition

$$\int_\Omega f d\mu = \sup_{g\in SF^+(f)} \int g(\omega) d\mu \quad \text{and} \quad \int_\Omega f_n d\mu = \sup_{g\in SF^+(f_n)} \int g(\omega) d\mu$$

and that the fact $f \geq f_n$ tells us that $SF^+(f_n) \subset SF^+(f)$, then we get the result just using the definition of supremum we get

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

$\underline{\int f d\mu \leq \lim_{n\to\infty} \int_\Omega f_n d\mu:}$

For any $s \in SF^+(f)$, define $\nu(S) := \int_S s d\mu$ for $S \in \mathcal{F}$. Now take $\alpha \in (0,1)$ and let

$$S_n := \{\omega : \alpha s(\omega) \leq f_n(\omega)\}.$$

Then S_n is measurable because it is the pre-image of $(\alpha s - f)^{-1}([0,\infty))$, since s and f are measurable. S_n is increasing with n as f_n is increasing. Moreover, not only are they increasing, their limit is the whole set since any $\omega \in \Omega$ belongs to S_n for sufficiently large n since ultimately $\alpha s \leq f$ and $f_n \uparrow f$.

That is, $S_1 \subset \dots$ and $\bigcup S_n = \Omega$. Hence

$$\int_\Omega s d\mu = \nu(\Omega) = \lim_{n\to\infty} \nu(S_n) = \lim_{n\to\infty} \int_{S_n} s d\mu$$

where the middle equality is because ν is a measure, hence continuous and can pass limits.

Since we know $\alpha s \leq f_n$ on S_n, and for simple functions we can check by definition that

$$\alpha \int_{S_n} s d\mu = \int_{S_n} \alpha s d\mu$$

so we have by above

$$\alpha \int_{S_n} s d\mu = \int_{S_n} \alpha s d\mu \leq \int_{S_n} f_n d\mu \leq \int_\Omega f_n d\mu$$

which is true for all $\alpha \in (0,1)$ and $s \in SF^+(f)$.

Now we conclude that since simple functions are well-defined on all measurable set

$$\int_\Omega s d\mu = \lim_{n\to\infty} \int_{S_n} s d\mu \leq \frac{1}{\alpha} \lim_{n\to\infty} \int_\Omega f_n d\mu$$

and by taking the sup on both sides

$$\int_\Omega f d\mu = \sup_{s\in SF^+(f)} \int_\Omega s d\mu \leq \sup_{s\in SF^+(f)} \frac{1}{\alpha} \lim_{n\to\infty} \int_\Omega f_n d\mu = \frac{1}{\alpha} \lim_{n\to\infty} \int_\Omega f_n d\mu$$

where the last part we take out sup since the bound is uniform in s. But then since α is arbitrary in $(0,1)$, thus taking off α is fine here. And thus we get

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

$\square$

Note that till here the non-negativity condition cannot be dropped.

Proposition 5.8. *Given any measurabel function f that is non-negative. Then there exists a sequence of non-negative simple functions that monotonically increases to f.*

Proof. (by construction)

For fixed n, the equation is always true

$$k2^{-n} \leq f(\omega) < (k+1)2^{-n}$$

so that we can break $f(\omega)$ into intervals of length 2^{-n}.

Define

$$f_n(\omega) = \begin{cases} k2^{-n} & \text{if } k \leq n2^n, i.e. f(\omega) \leq n \\ n & \text{elsewhere} \end{cases}$$

then we can check that f is increasing and the error we make goes to 0. □

Now we'll leak some of the homework solutions just because they are so important.

Proposition 5.9. *(Change of variables, proof statement) For the measure space $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$. Let $g : \Omega_1 \to \Omega_2$ and $f : \Omega_2 \to \mathbb{R}$ be measurable functions such that $f \circ g$ is integrable.*

We can let ν be the measure on Ω_2 induced by g:

$$\nu(A) := \mu_1(g^{-1}(A)).$$

Then

$$\int_{\Omega_1} f \circ g d\mu_1 = \int_{\Omega_2} f d\nu.$$

Proposition 5.10. *(Linearity of integral) If f, g are 2 integrable functions, then for any $\alpha, \beta \in \mathbb{R}$, the function $\alpha f + \beta g$ is integrable. Moreover we can explicitly write*

$$\int_{\Omega} \alpha f + \beta g d\mu = \alpha \int_{\Omega} f d\mu + \beta \int_{\Omega} g d\mu.$$

If f, g are non-negative, then integrability is not needed for the case of $\infty + \infty$.

Proof. We prove for the case when f, g are non-negative. For simple functions, the result is obvious since it's just finite sums. By construction in proposition 5.3, we can assume that $v_n \uparrow g, u_n \uparrow f$ such that

$$\int (f+g) d\mu \overset{MCT}{=} \lim_{n\to\infty} \int (u_n + v_n) d\mu = \lim_{n\to\infty} \left(\int u_n d\mu + \int v_n d\mu \right) \overset{MCT}{=} \int f d\mu + \int g d\mu.$$

□

Theorem 5.11. *(Fatou's Lemma) Let $\{f_n\}$ be a sequence of non-negative measurable functions, then*

$$\int \liminf f_n d\mu \leq \liminf \int f_n d\mu.$$

We'll see that it's actually an easy result of the monotone convergence theorem.

Proof. Let $g_n := \inf_{m\geq n} f_m$, $g := \liminf f_n$. Then we know that g_n increases to g point wise. Also, since $g_n \leq f_n$ every where and g_n are measurable, we have

$$\int \liminf f_n d\mu = \int g d\mu \overset{MCT}{=} \lim_{n\to\infty} \int g_n d\mu \leq \liminf \int f_n d\mu$$

where the last inequality is because $g_n \leq f_m$ for all $m \geq n$ and hence

$$\int g_n d\mu = \int \inf_{m\geq n} f_m d\mu \leq \inf_{m\geq n} \int f_n d\mu$$

and $\liminf$ does not care about the first finite terms. □

We state the most important theorem till now.

Theorem 5.12. *(Dominated convergence theorem):*

Let $\{f_n\}$ be measurable such that $f_n \to f$ point wise. Assume $\exists h$ integrable such that $\forall n$, $|f_n| \leq h$ everywhere, then

$$\int |f_n - f| d\mu \to 0$$

in particular

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

The sketch of proof is the following: f is measurable since it is the limit of measurable functions. Since $|f| \leq h$, we know that $f + h \geq 0$ is non-negative and integrable. Then we can start using what we have from here.

Proof. (from next class) Let f_n be measurable such that $f_n \to f$, then f is measurable. Now since h measurable we know that $f_n + h$ and $f + h$ are all measurable functions. By Fatou's lemma we have that

$$\int (f + h) d\mu = \int \liminf (f_n + h) d\mu \leq \liminf \int (f_n + h) d\mu$$

which gives us that

$$\int f d\mu \leq \liminf \int f_n d\mu$$

and if we replace $f \to -f$, $f_n \to -f_n$ we have the opposite inequality

$$\int -f d\mu \leq -\liminf \int f_n d\mu$$

$$\Rightarrow \int f\,d\mu \geq \limsup \int f_n d\mu.$$

Combining these we get

$$\limsup \int f_n d\mu \leq \int f\,d\mu \leq \liminf \int f_n d\mu$$

then by definition of $\limsup$ and $\liminf$, we get

$$\int f\,d\mu = \lim_{n\to\infty} \int f\,d\mu.$$

$\square$

It's left an exercise to prove that

$$\left|\int_\Omega f\,d\mu\right| \leq \int_\Omega |f|\,d\mu.$$

6. 1/25: PRODUCT MEASURES; PROBABILITY RECAP

We finish up with (no proof) of the product measures.

6.1. Product measures.

Def 6.1. *Let $(X_i, \mathcal{A}_i)$ be a collection of measurable spaces (where finite and infinite does make a difference, but we will delay the part of infinite to next semester). Now let $X_1 \times \cdots \times X_n$ be the product space, then the σ-algebra generated by all rectangles is the product σ-algebra, where rectangles are sets of the following form:*

$$A_1 \times \cdots \times A_n$$

for $A_i \in \mathcal{A}_i$.

Proposition 6.2. *Let $(X, \mathcal{A})$ be as above and each X_i equipped with measure μ_i, then $\exists \mu$ on X such that μ on any $A \in \mathcal{A}$ has that*

$$\mu(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \mu_i(A_i).$$

Proof. We can use the extension theorem by checking that the collection is a boolean algebra and that μ is a premeasure. Then prove uniqueness, but that's the easy part. do it. □

Below is a lemma that will be used in the proof of Fubini's theorem.

Lemma 6.3. *Let $(X_i, \mathcal{A}_i)$ with $i = 1, 2, 3$ be 3 measurable spaces with*

$$f : X_1 \times X_2 \to X_3$$

being measurable. Then for each $x \in X_1$

$$y \mapsto f(x, y)$$

is measurable on x_2.

Well, we skip the proof.

Theorem 6.4. *(Fubini-Tonelli Theorem):*

(i) We have that 2 σ- finite measure spaces $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$. Let $\mu = \mu_1 \times \mu_2$, $f : X_1 \times X_2 \to [0, \infty]$, $A_1 \times A_2$ be measurable. Then, if f is either integrable or non-negative

$$\int_{X_1 \times X_2} f \, d\mu = \int_{X_1} \int_{X_2} f \, d\mu_2 d\mu_1 = \int_{X_2} \int_{X_1} f \, d\mu_1 d\mu_2$$

which is well-defined since we are integrating $f(x, y)$ for fixed $x \in X_1$, which by lemma is measurable.

(ii) If f is integrable for μ_1, then for almost every x (the complement's measure is 0), we have
- $y \mapsto f(x, y)$ *is* μ_2 *integrable;*
- $x \mapsto \displaystyle\int_{X_2} f(x, y) d\mu_2$ *is integrable.*

This is indeed a very abrupt ending of our measure part of the course, but let's start probability now.

6.2. **Probability.**

We begin with some basic definitions.

Def 6.5.

- *The base space Ω we are looking is called in various ways:universe; ambient space; sample space; outcome space.*
- *Ω is equipped with a σ-algebra $\mathcal{F}$ and the elements of $\mathcal{F}$ are called events.*
- *The sample space is equipped with measure $\mathbb{P}$ such that $\mathbb{P}(A) \in [0, 1]$, which is "the likelihood of the corresponding event." We denote it by probability.*

Proposition 6.6. *The probability measure satisfies*

- $\mathbb{P}(A \text{ or } B) = \mathbb{P}(A \cup B)$ *and* $\mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B)$*;*
- $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$.

Now we introduce the dictionary that we will use in probability:

Remark 6.7.

(1) A measurable function $f : \Omega \to \mathbb{R}$ or $\mathbb{R}^*$ is called a random variable, i.e. a measurement of experiments.

(2) Given random variables $X, Y : \Omega \to \mathbb{R}$, then
   - $\mathbb{P}(\omega \in \Omega : X(\omega) \in A) = \mathbb{P}(X \in A)$ for $A \subset \mathbb{R}$;
   - $\mathbb{P}(\omega \in \Omega : X(\omega) \in A, Y \in B) = \mathbb{P}(X \in A, Y \in B)$.

(3) Given a measurable $f : \mathbb{R} \mapsto \mathbb{R}$, then we denote

$$f \circ X = f(X).$$

(4) For the probability measure $\mathbb{P}$, its integration is called the expectation

$$\int_\Omega X d\mathbb{P} = \mathbb{E}[X]$$

which works sort of like the average of measurement. This is actually justified through LLN.

(5) Event E is said to hold almost surely if $\mathbb{P}(E) = 1$.

(6) A σ-algebra generated by the set

$$\{\omega \in \Omega : X(\omega) \in \mathcal{A}\}$$

for all $\mathcal{A}$ Borel, is $\sigma(X)$.

(7) For random variable $X : \Omega \to \mathbb{R}$, we can define the law of distribution of X by

$$\mu_X(A) := \mathbb{P}(X \in A) := \mathbb{P}(\{\omega \in \Omega | X(\omega) \in A\}).$$

Note we often don't distinguish X with μ_X, and it's very common to work directly with the law of a random variable, rather than the variable itself.

Def 6.8. *The function $t \mapsto F_X(t) = P(X \leq t)$ is the cumulative distribution function of X.*

Proposition 6.9. *Let $F : \mathbb{R} \to [0, 1]$ be a function that is non-decreasing, right continuous, and that*

$$\lim_{t\to-\infty} F(t) = 0, \lim_{t\to\infty} F(t) = 1$$

then there exists space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variable X such that $F_X(t) = F(t)$. Conversely, any CDF satisfy those properties.

Proof. We check the CDF is non-decreasing by monotonicity of the product measure, it is right continuous by the continuity of measure (downward limit).

Now, we prove the forward direction, which is less trivial. We construct the random variable X on $((0, 1), \mathcal{B}(0, 1), \mathcal{L}(0, 1))$ by

$$X(\omega) := \inf\{y \in \mathbb{R} : F(y) \geq \omega\}$$

which is measurable since check right continuous functions are measurable, which means that it is indeed a random variable.

We claim that

$$\{\omega : X(\omega) \leq t\} = \{\omega : F(t) \geq \omega\}.$$

If this is true, then we have that

$$F(t) = \lambda\{\omega : F(t) \geq \omega\} = \lambda\{\omega : X(\omega) \leq t\} = F_X(t)$$

where the last equality is by definition of CDF, and λ is Lebesgue measure (the measure we've defined it on).

So we will be done if we can show that claim. We do it in both directions.

$\supset$:

If $\omega \leq F(t)$, then $X(\omega) \leq t$ by definition, since $t \in \{y \in \mathbb{R}, F(y) \geq \omega\}$ and $X(\omega)$ is the inf of that class.

$\subset$:

Suppose $X(\omega) \leq t$. For contradiction, assume that $F(t) < \omega$, then $F(t+\varepsilon) < \omega$ for small ε. This, however, contradicts with the fact that $X(\omega) \leq t$ since $X(\omega) = \inf\{y \in \mathbb{R} : F(y) \geq \omega\}$, the inf of y such that $F(y) \geq \omega$. yet it sure is not the smallest since all $t + \varepsilon$ satisfies the condition. Contradiction.

□

Proposition 6.10. *2 random variables have the same CDF iff they have the same law.*

Proof. The backwards direction is just by definition of CDF.

For the forward direction, we use our $\pi - \lambda$ theorem on the family of sets $(-\infty, t)$, which is a π system. Then since $\sigma((-\infty, t)) = \mathcal{B}$ we are done by the theorem. □

Def 6.11. *If the image of X is discrete, i.e. finite or countable, then the random variable X is called discrete.*

Def 6.12. *Let f be a non-negative integrable function on $\mathbb{R}$ such that*

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

where dx is the Lebesgue measure. We say that the random variable X has a probability density function f_X iff the CDF

$$F_X(t) = \int_{-\infty}^{t} f_X(y)dy, \ \forall t \in \mathbb{R}$$

then $f(x)$ is called the density of X and X is a continuous random variable.

7. 2/1: MORE ON PROBABILITY; INDEPENDENCE

7.1. **More on continuous random variables.**

Def 7.1. *Let f be a non-negative function on $\mathbb{R}$ with*

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

If we define a probability measure

$$\nu(A) := \int_A f(x)dx$$

then f is the density for measure ν.

Moreover, if ν is the law of a random variable X which is absolutely continuous, we call f the density of X.

Proposition 7.2. *For random variable X, we can check that such f is its density only on intervals $[a,b]$ for all $a,b \in \mathbb{R}$ and that a,b are continuous points of F_x.*

The way to do this is partially homework problem. First, let F_x be the cdf of X. Then, it has at most countable number of discontinuity points (since right continuous).

Now, for any a,b, the continuity points of F, we have that

$$\nu(A) := \int_a^b f(x)dx.$$

For other $\{a,b\}$, we use the homework exercise and the dominated convergence theorem to prove the same equality. Now, using $\pi - \lambda$ theorem, where the π system is all intervals, and λ system are all Borel sets, then the equality holds everywhere.

7.2. **Variance and covariance.**

Def 7.3.

- *The* <u>*variance*</u> *of a random variable is*

  $$\mathrm{Var}(X) := \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

  given that the value is finite.
- *The* <u>*standard deviation*</u> *is $\sigma = \sqrt{\mathrm{Var}(X)}$.*
- *$\mathbb{E}[X^k]$ is the k-th moment.*

Proposition 7.4. *(Chebyshev's inequality) Assume $\mathbb{E}[X^2] < \infty$, then the following holds: for all $t \geq 0$*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Whereas to prove this we just use Markov's inequality:

Proposition 7.5. *(Markov's inequality)For $X \geq 0$*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

And to prove Chebyshev we just take $Y = |X - \mathbb{E}[X]|$ and take the square on both sides.

Def 7.6. *The moment generating function is*

$$m_X(t) := \mathbb{E}[e^{tX}]$$

for $t \in \mathbb{R}$, provided it exists.

Why is it the "moment generating function"? Well, we look at the pseudo-expansion and get a clue:

$$e^{tX} \approx 1 + tX + \frac{t^2X^2}{2} + \dots$$

and take the expectation.

So one way we might approximate is to use the above and get

$$\mathbb{P}(X \geq t) = \mathbb{P}(e^{sX} \geq e^{st}) \leq \frac{\mathbb{E}[e^{sX}]}{e^{st}}.$$

Def 7.7. *The covariance is*

$$\operatorname{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Def 7.8. *The correlation is*

$$\operatorname{Cor}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\operatorname{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

7.3. **Independence.**

We gradually build up what is independence.

Def 7.9. *2 events are called independent if*

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

Remark 7.10. *This is a notion of independence because by the Bayesian formula we have*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

which is intuitive. Now, given the formula for independence we have that $\mathbb{P}(A) = \mathbb{P}(A|B)$, so this means that the probability of A doesn't care about the existence of B.

Def 7.11. *2 σ-algebra $\mathcal{H}, \mathcal{G} \subset \mathcal{F}$ are independent if $\forall H \in \mathcal{H}, G \in \mathcal{G}$, we have H and G are independent.*

Def 7.12. *Let X and Y be two random variables. Then they are independent if $\sigma(X)$ and $\sigma(Y)$ are independent.*

Lemma 7.13. *If A, B are independent, thhen so are A^c, B^c, A, B^c, and A^c, B.*

Proof.
$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(B)\mathbb{P}(A^c).$$
and the rest pairs are proven similarly. □

Def 7.14. *A collection of event is mutually independent if for $A_i \in \mathcal{F}$ and any $L < \infty$, we have*
$$\mathbb{P}(A_{i,1} \cap \cdots \cap A_{i,L}) = \prod_{j=1}^{L} \mathbb{P}(A_{i,j}).$$

Note that pairwise independent is a strictly weaker condition.

Theorem 7.15. *Let $\mathcal{A}_i \subset \mathcal{F}$ be a collection of sets in $\mathcal{F}$. Then, let $G_i = \sigma(\mathcal{A}_i)$ for $i = 1, \ldots, n$. Assume that $\mathcal{A}_i$ are π-systems, then a sufficient condition for G_i to be mutually independent is that $\mathcal{A}_i$ are mutually independent.*

Corollary 7.16. *In order for random variable $X_1, \ldots, X_n$ to be independent, it is sufficient to show that*
$$\mathbb{P}(X_1 \le t_1, X_2 \le t_2, \ldots, X_n \le t_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \le t_i)$$
where note the right hand side is the cdf for X_i.

Proof. (Theorem 7.15):

We use induction to prove this. For the base case with only G_1 and G_2, we only need to check that, for $A_1 \in G_1, A_2 \in G_2$
$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).$$

But for this step we can use $\mathbb{P}$ to be the first measure, and use
$$\nu(A_1) := \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_2)}$$
as the second measure. Then since these measures coincides for all $A_i \in \mathcal{A}_i$, we can extend the equality to $A_i \in \mathcal{G}_i$, and hence concludes the proof for the base case.

Now, for more than 2 sets, we first assume that $G_1, \ldots, G_{n-1}$ are mutually independent, and let $H = A_1 \cap \cdots \cap A_{n-1}$ for $A_i \in \mathcal{A}_i$. Then, pick $A \in \mathcal{A}_n$. Again we can define
$$\mu(A) = \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)}$$
and use π-λ theorem to extend the measure, hence get the result. □

Now we start to consider what is independent for a measurable function f.

Theorem 7.17. *If for $1 \leq i \leq n, 1 \leq j \leq m(i)$ we have X_{ij} are independent and that $f_i : \mathbb{R}^{m(i)} \to \mathbb{R}$ are measurable. Then we know that*

$$f_i(X_{i,1}, \dots, X_{i,m(i)})$$

are independent in the sense that a collection of random variables are independent.

Theorem 7.18. *Suppose $X_1, \dots, X_n$ are independent random variables and X_i has laws μ_i, then $(X_1, \dots, X_n)$ has the law $\mu := \mu_1 \times \mu_2 \times \dots \times \mu_n$, and*

$$\mu(A_1 \times A_2 \times \dots A_n) = \prod \mu_i(A_i).$$

Proof. By independence, we have

$$\begin{aligned} \mu(A_1 \times A_2 \times \dots A_n) &= \mathbb{P}\big((X_1, \dots, X_n) \in A_1 \times \dots \times A_n\big) \\ &\overset{indep}{=} \prod_{i=1}^{n} \mathbb{P}(x_i \in A_i) \overset{def}{=} \prod_{i=1}^{n} \mu_i(A_i) \end{aligned}$$

Now, since the rectangles generates a π system, and we have σ finiteness, the equality holds on the extension. □

Corollary 7.19. *Combine the above result with the f_i result we have that*

$$m_{X+Y}(t) = m_X(t)m_Y(t).$$

Proof.

$$m_{X+Y}(t) = \mathbb{E}\left[e^{t(X+Y)}\right] = \mathbb{E}\left[e^{tX}e^{tY}\right] = \mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right] = m_X(t)m_Y(t).$$

□

Moreover, the result is proven in the same way for the characteristic function (well, of course).

Def 7.20. *The characteristic function of random variable X is*

$$\phi_X(t) = \mathbb{E}\left[e^{itX}\right]\bigg(= \mathbb{E}\left[\cos tx + i \sin tx\right].\bigg)$$

And for a similar reason we have

Corollary 7.21. *If X, Y are independent, then*

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

7.4. **multivariate normal distribution.** :

Def 7.22. *A <u>random vector</u> is a function from* $(X, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^n$.

Proposition 7.23. *If for all vector a, we know that* $a \cdot X$ *is normal, then we can equivalently say that*

$$X = DW + \mu$$

where D *is an* $n \times n$ *vector,* W *is a vector whose entries are independent normal, and* μ *is a scaler vector.*

Def 7.24. *For non-degenerate normal random vector, there exists an explicit formula for the density of their joined cdf.*

8. 2/6: Lemmas for preparation; Notions of convergence

8.1. **Construction of measure and random variables.**

We start with a fun fact of how to construct the Lebesgue measure on $[0, 1]$, i.e. the measure space $((0, 1), \mathcal{B}(0, 1), \lambda)$.

To do this we consider the decimal expression of any number between 0 and 1, and we let $0.999999\dots$ to be of the same representation as $1.000\dots$.

Now we will construct the sequence of random variable $X_n(\omega) = \omega_n$, the n-th digit of ω. Thus, from definition $\omega = \sum_{n\geq 1} \frac{X_n(\omega)}{10^n}$.

Then, if we consider just iid uniformly distributed random variable Y_n on $\{0, \dots, 9\}$, then define

$$Y = \sum_{n\geq 1} \frac{Y_n(\omega)}{10^n}$$

then the law of Y is precisely the Lebesgue measure on $[0, 1]$.

We've seen above how to construct a measure out of simple uniform random variables, but we might continue ask how to construct a sequence of independent random variables with given laws.

Let μ_i be some Borel probability measure on $\mathbb{R}$. We will construct a random variable r_i on $((0, 1), \mathcal{B}(0, 1), \lambda)$ whose law is μ_i by the following method:

Let μ_i be the law of Y_i, then the distribution function is such that $F_{Y_i}(t) = \mu_i((-\infty, t])$.

Then, define π_i as the projection from the product space $\Omega = \prod \Omega_i$ to Ω_i. Then define $X_i(\omega) = Y_i \circ \pi_i$ we can claim that X_i has the same law as Y_i, that is, it also has law μ_i. Since π_i are independent so are X_i.

8.2. **Toward limit theorems.**

The road map from now on is as follows: we are interested in the universality type of behavior, that is, LLN, CLT, etc. In particular, given X_i iid random variables and $\mathbb{E}[X_i] = \mu$, we can show things like

$$\frac{1}{N} \sum X_i \xrightarrow{p} \mu$$

or

$$\frac{\sum (X_i - \mu)}{\sigma \sqrt{n}} \xrightarrow{d} N(0, 1)$$

under certain conditions. But we need to define some notions of convergence, prove some preliminary lemmas to proceed.

Def 8.1. *For a sequence of subsets $A_n \subset \Omega$, define*

$$\begin{aligned}\limsup_{n\to\infty} A_n :&= \bigcap_{n=1}^{\infty}\bigcup_{l=n}^{\infty} A_l = \left\{ w \middle| w \in A_n \text{ for infinitely many } A_n \right\} \\ &= \{A_n \ \text{i.o. (infinitely often)}\}.\end{aligned}$$

Lemma 8.2. *(Borel-Cantelli first lemma) On the measure space $(\Omega, \mathcal{F}, \mathbb{P})$, suppose for some sequence $A_n \subset \Omega$ we have*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$$

then

$$\mathbb{P}(A_n \ i.o.) = \mathbb{P}(\limsup_{n\to\infty} A_n) = 0.$$

Note that the converse is not true since we can just pick any cover that has empty intersection.

Proof. Define

$$N(\omega) = \sum_{k=1}^{\infty} \mathbb{1}_{A_k}(\omega)$$

then

$$N = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{1}_{A_k}(\omega) =: \lim_{n\to\infty} S_n$$

and

$$\mathbb{E}\left[\lim_{n\to\infty} S_n\right] = \lim_{n\to\infty} \sum_{n=1}^{n} \mathbb{P}(A_n) \nearrow \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$$

so S_n is bounded above by an integrable function, thus by MCT we can exchange the limit and get

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{k=1}^{n} \mathbb{1}_{A_k}\right] = \mathbb{E}[N] = \sum_{i=1}^{\infty} \mathbb{P}(A_n) < \infty.$$

But then this means

$$\mathbb{P}(A_n \ i.o.) = \mathbb{P}(\omega : N(\omega) = \infty) = 0$$

and we are done. □

Lemma 8.3. *(Second Borel-Cantelli Lemma) If $\sum_{n\geq 1} \mathbb{P}(A_n) = \infty$ and A_n are mutually independent, then*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 1.$$

We'll skip the proof however.

After seeing the Borel-Cantelli lemmas, let's look at some 0-1 laws.

Def 8.4. *Let $X_1, X_2, \dots$ be a sequence of random variables defined on the same probability space, then the tail σ-algebra is*

$$\mathcal{T}(X_1, \dots) = \bigcap_{n=1}^{\infty} \sigma(X_n, \dots)$$

or we can think of the right hand side as the σ-algebra generated by events $X_i((-\infty, t])$ for any $t \in \mathbb{R}$, $i = n, n+1, \dots$.

Theorem 8.5. *(Kolmogorov's 0-1 law) Let $\{X_n\}$ be independent family of mutually independent random variables. Then for any $A \in \mathcal{T}$, $\mathbb{P}(A) = 0$ or 1. In other words, the tail σ-algebra is trivial.*

Corollary 8.6. *A random variable is measurable with respect to the tail σ-algebra if it is constant almost everywhere.*

Proof. (Corollary 8.6) Consider the cdf of X

$$F_X(t) = \mathbb{P}((-\infty, t])$$

since the random variable is measurable with respect to the tail, it means $X^{-1}(-\infty, t]$ is in the tail σ-algebra, so the measure of it is either 0 or 1. But this immediately implies that $x = t := \sup\{t | F_X(t) = 0\}$ almost everywhere. $\square$

Proof. (Theorem 8.5)

The big idea is to show that $A \in \mathcal{T}$ are independent from themselves. If so, then we have

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$$

which means $\mathbb{P}(A) = 0$ or 1.

Let $A \in \mathcal{T}$. For any n, consider the event $B \in \sigma(X_1, \dots, X_n)$, then A is independent from B since B must be in the σ-algebra of some X_k and A is in the σ-algebra of X_{n+1}, plus X_i are independent.

Take

$$F_\infty := \bigcup_{n \geq 1} \sigma(X_1, \dots, X_n)$$

then for all $B \in \mathcal{F}_\infty$, $B \in \sigma(X_1, \dots, X_n)$ for some n, thus A is independent of B.

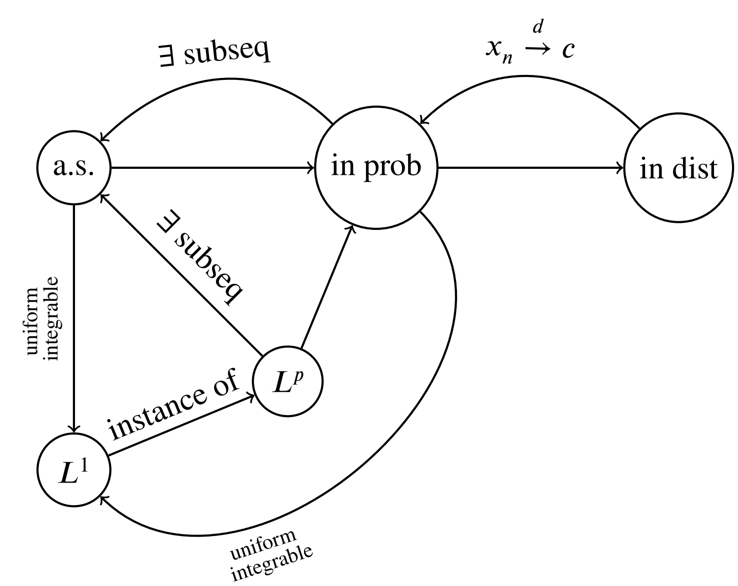
But then note that $\mathcal{T}$ lies inside the σ-algebra genreated by $\mathcal{F}_\infty$, so by π-λ theorem A is independent from itself. $\square$

8.3. **Four notions of convergence.**

Def 8.7.

- *A sequence of random variables $\{X_n\}$ is said to converge a.s. (almost surely) to X if*
$$\lim_{n\to\infty} X_n = X$$
a.s., or equivalently
$$\mathbb{P}\left(\omega \middle| \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\right) = 0.$$
- *A sequence $\{X_n\} \to X$ in probability if for any $\varepsilon > 0$*
$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$
- *For each n, let F_{X_n} be the cdf of X_n. Then X_n is said to converge in distribution if*
$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$$
for any t that is a continuity point of F_X.
- *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space with $1 \leq p < \infty$ we define the L^p norm as*
$$||f||_{L^p} := \left(\int |f|^p d\mu\right)^{\frac{1}{p}}$$
from which we can say that X_n converge in L^p to X if
$$\lim_{n\to\infty} ||f_n - f||_p = 0.$$
- *When $X_n \to X$ in L^1, we also call it convergence in expectation.*

Now we give an overview of the relations between these convergences:



Proposition 8.8. *Convergence in prob $\Rightarrow$ convergence in distribution.*

Proof. We need to show that $\mathbb{P}(X_n \le t) \to \mathbb{P}(X \le t)$. fill in. $\square$

Note that the above statement is not true if we require the limit to be passed for all point of F_X (not just the continuity points).

Proposition 8.9. *Convergence a.s. $\Rightarrow$ convergence in probability.*

Proof. fill in. $\square$

Proposition 8.10. *Convergence in L^p $\Rightarrow$ convergence in probability.*

Proof. Pick $\varepsilon > 0$, then we have by Markov's inequality that

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p \ge \varepsilon^p) \le \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p} \to 0$$

and we are done. $\square$

Note that in the graph of relations we mentioned uniform integrable, which we define now.

Def 8.11. *A sequence of random variable is <u>uniformly integrable</u> if for any $\varepsilon > 0$, there $\exists k > 0$ such that $\forall n$ (uniform in n) we have*

$$\mathbb{E}\left[|X_n| \cdot \mathbb{1}_{|X_n|>k}\right] \le \varepsilon.$$

Example 8.12. *Let's see two examples here.*

(1) If $\{X_n\}$ is dominated, i.e. $\exists Y \ge 0$ integrable and $|X_n| \le Y$ a.s. for all n. Then X_n is uniformly integrable since

$$\mathbb{E}\left[|X_n| \cdot \mathbb{1}_{|X_n|>k}\right] \le \mathbb{E}[Y \cdot \mathbb{1}_{Y>k}] \xrightarrow{k\to\infty} 0.$$

(2) When $\{X_n\}$ is bounded in L^p for $1 \le p < \infty$, then the sequence is uniformly integrable. Since by contradiction the proof is clear.

9. 2/8: LLN

Now we give an equivalent definition of convergence in distribution:

Proposition 9.1. *A sequence of random variables X_n converges in distribution $\iff$ for all bounded continuous function f*

$$\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X)\right].$$

Corollary 9.2. *If $X_n \xrightarrow{d} X$, then $f(X_n) \xrightarrow{d} f(X)$ for bounded continuous f.*

Proof. Now we prove the easier direction of proposition 9.1. The other direction is in book.

$\Leftarrow$:

Assume that $\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X)\right]$ for all bounded continuous f. Then we can choose

$$f(y) = \begin{cases} 1 & y \leq t \\ 0 & y \geq t+\varepsilon \end{cases}$$

and let f be continuously decreasing in between. Thus we have the following inequalities

$$\mathbb{E}[f(X)] = \int_{-\infty}^{t} X dx + \int_{t}^{t+\varepsilon} f(X) dx \leq \int_{-\infty}^{t+\varepsilon} X_n dx = F_X(t+\varepsilon)$$

$$\mathbb{E}\left[f(X_n)\right] = \int_{-\infty}^{t} X_n dx + \int_{t}^{t+\varepsilon} f(X_n) dx \geq \int_{-\infty}^{t} X_n dx = F_{X_n}(t)$$

and thus

$$\limsup_{n\to\infty} F_{X_n}(t) \leq \limsup_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \leq F_X(t+\varepsilon).$$

Similarly, we can choose another function

$$g(y) = \begin{cases} 1 & y \leq t-\varepsilon \\ 0 & y \geq t \end{cases}$$

and let g be continuously decreasing in between. Then with a similar argument we have

$$\liminf_{n\to\infty} F_{X_n}(t) \geq F_X(t-\varepsilon).$$

At this point we see that if t is a continuous point of F_X (crucial condition in definition!), then we can take $\varepsilon \to 0$ and since the left limit and right limit of $F_X(t)$ coincides we get

$$\liminf_{n\to\infty} F_{X_n}(t) \geq F_X(t) \geq \limsup_{n\to\infty} F_{X_n}(t)$$

and hence

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t).$$

□

Now we start to prove the law of large numbers, which in general is a statement that given some conditions we have

$$\frac{1}{N}\sum X_i \to \mathbb{E}[X] =: \mu$$

in different sense of convergence.

Theorem 9.3. *(Weak LLN) Let* $\{X_i\}$ *be uncorrelated and suppose* $\mathrm{Var}(X_i) \le C$, $\mathbb{E}[X_i] = \mu$. *Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{L^2} \mu.$$

Proof. First, we denote $S_n := \sum_{i=1}^{n} X_i$.

It suffices to show that

$$\mathbb{E}\left[\left(\frac{1}{n}S_n - \mu\right)^2\right] = 0$$

but note that $\mathbb{E}\left[\frac{1}{n}S_n\right] = \mu$ we have

$$\mathbb{E}\left[\left(\frac{1}{n}S_n - \mu\right)^2\right] = \mathrm{Var}\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}\mathrm{Var}(S_n))$$

and since correlation is 0 means

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$$

(easy calculation) and thus

$$\mathbb{E}\left[\left(\frac{1}{n}S_n - \mu\right)^2\right] \le \frac{Cn}{n^2} \to 0$$

and we are done. $\square$

Remark 9.4. *Now we note that it really is the same as convergence in probability, since*

$$\mathbb{P}\left(\left|\frac{1}{n}S_n - \mu\right| \ge \varepsilon\right) = 0$$

implies

$$\int \left(\left|\frac{1}{n}S_n - \mu\right|\right)^2 d\mu \le \varepsilon^2$$

and the other direction is by the relation graph.

Now we look at strong law of large numbers. Rather than giving a complete form of the theorem we first look at a deformed version.

Def 9.5. *We define the* <u>*truncation*</u> $\overline{X}$ *of a random variable* X *such that*

$$\overline{X} := X \cdot \mathbb{1}_{|X| \leq M} = \begin{cases} X & |X| \leq M \\ 0 & otherwise \end{cases}.$$

Note that this is a general notation and we should specify the M *as we use the notation.*

Theorem 9.6. *For each* n*, let* $X_{n,k}$ *be mutually independent variables with* $1 \leq k \leq n$*. Moreover, let* b_n *be a positive real sequence that goes to* ∞*. We truncate by letting*

$$\overline{X_{n,k}} = X_{n,k} \cdot \mathbb{1}_{\{|X_{n,k}| \leq b_n\}}.$$

Assume that

(1) $\displaystyle\sum_{k=1}^{n} \mathbb{P}(|X_{n,k}| > b_n) \to 0$ *as* $n \to \infty$*;*

(2) $\displaystyle\frac{1}{b_n^2} \sum_{k=1}^{n} \mathbb{E}\left[\overline{X}_{n,k}^2\right] \to 0$ *as* $n \to \infty$*.*

Then

$$\frac{S_n - a_n}{b_n} \xrightarrow{p} 0$$

for $S_n := \displaystyle\sum_{k=1}^{n} X_{n,k}$ *and* $a_n := \displaystyle\sum_{k=1}^{n} \mathbb{E}\left[\overline{X_{n,k}}\right]$*.*

The theorem look bizarre indeed. We'll see that the assumptions actually comes from what we exactly need in the proof. One way we can visualize this theorem is to view $X_{n,k}$ as a triangle of variables:

$$\begin{matrix} X_{1,1} & & & \\ X_{2,1} & X_{2,2} & & \\ X_{3,1} & X_{3,2} & X_{3,3} & \\ \vdots & \vdots & & \ddots \end{matrix}$$

and S_n as the sum over the n-th row.

Proof. First fix $\varepsilon > 0$.

What we want to show is that

$$\mathbb{P}\left(\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon\right) \to 0$$

so let's try to bound this probability with our assumptions:

$$\mathbb{P}\left(\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon\right) \leq \mathbb{P}\left(\left|\frac{\overline{S_n} - a_n}{b_n}\right| > \varepsilon\right) + \mathbb{P}\left(S_n \neq \overline{S_n}\right)$$

since one of them holds. Then we divide and conquer by

$$\mathbb{P}\left(S_n \neq \overline{S_n}\right) \leq \mathbb{P}\left(\bigcup_{k=1}^{n}\left\{\overline{X_{n,k}} \neq X_{n,k}\right\}\right) \leq \sum_{k=1}^{n}\mathbb{P}(|X_{n,k}| > b_n) \to 0$$

by assumption (1) and

$$\begin{aligned}\mathbb{P}\left(\left|\frac{\overline{S_n} - a_n}{b_n}\right| > \varepsilon\right) &\overset{Chebyshev}{\leq} \frac{1}{\varepsilon^2}\mathbb{E}\left[\left(\frac{\overline{S_n} - a_n}{b_n}\right)^2\right] \leq \frac{1}{\varepsilon^2 b_n^2}\operatorname{Var}\left(\overline{S_n}\right) \\ &\overset{independence}{=} \frac{1}{\varepsilon^2 b_n^2}\sum_{i=1}^{n}\operatorname{Var}\left(\overline{X_{n,k}}\right) = \frac{1}{\varepsilon^2 b_n^2}\sum_{i=1}^{n}\left(\mathbb{E}\left[\overline{X}_{n,k}^2\right] - \mathbb{E}\left[\overline{X}_{n,k}\right]^2\right) \\ &\leq \frac{1}{\varepsilon^2 b_n^2}\sum_{i=1}^{n}\mathbb{E}\left[\overline{X}_{n,k}^2\right] \to 0\end{aligned}$$

for ε fixed by assumption (2). Hence we are done. $\square$

This argument seems a little bit fishy since all we did is nothing but assuming what we need. Now we see how another form of LLN can be proven with this.

Theorem 9.7. *Let $\{X_n\}$ be a sequence of iid random variable with*

$$x \cdot \mathbb{P}(|X_i| > x) \to 0$$

as $x \to \infty$, and

$$S_n := \sum_{i=1}^{n} X_i, \quad \mu_n = \mathbb{E}\left[X_n \cdot \mathbb{1}_{\{|X_n|<n\}}\right].$$

Then

$$\left(\frac{1}{n}S_n - \mu_n\right) \xrightarrow{p} 0.$$

Proof. We note that we just take $X_{n,k} := X_k$ and $b_n = n$, that is the triangle of variables becomes

$$\begin{array}{llll} X_1 & & & \\ X_1 & X_2 & & \\ X_1 & X_2 & X_3 & \\ \vdots & \vdots & & \ddots \end{array}$$

and check that all condition of theorem 9.6 is satisfied. (See book 2.2.12). $\square$

With a few remarks we head into strong LLN:

Remark 9.8.

(1) Assume $\mathbb{E}[X_i] < \infty$, then

$$x \cdot \mathbb{P}(|X_1| \geq x) \leq \cdots \leq \mathbb{E}\left[|X_1| \cdot \mathbb{1}_{\{|X_1|>k\}}\right] \to 0$$

where the last inequality is by assumption and the thing in middle is integral identities for expectation.

(2) Note that the above assumption $\mathbb{E}[X_i] < \infty$ is strictly stronger than the result $x \cdot \mathbb{P}(|X_1| \geq x) \to 0$, since there exists example of random variables with that condition whose first moment is infinite.

(3) Assume X_1 non-negative with $x \cdot \mathbb{P}(X_1 \geq x) \to 0$, $x \to \infty$, then $\mathbb{E}\left[|X_1|^{1-\varepsilon}\right] < \infty$.

Proposition 9.9. *(Integral Identity)*

$$\mathbb{E}[Y^p] = \int py^{p-1}\mathbb{P}(Y \geq y)dy.$$

Thus, if we let $p = 1 - \varepsilon$, then we have

$$\mathbb{E}[Y^{1-\varepsilon}] = \int (1-\varepsilon)y^{-\varepsilon-1}y\mathbb{P}(Y \geq y)dy < \infty$$

by p-test and the condition that $y\mathbb{P}(Y \geq y)$ goes to 0.

10. 2/13: PROOF OF STRONG LLN; CHARACTER FUNCTION

Theorem 10.1. *(Strong LLN) (Etemadi) Let $\{X_n\}$ be pairwise independent identically distributed random variables with*

$$\mathbb{E}[|X_1|] < \infty$$

then

$$\frac{1}{n}\sum_{i=1}^{\infty} X_i \xrightarrow{as} \mathbb{E}[X_1].$$

We first introduce a lemma that we'll use during the proof.

Lemma 10.2. *For any $\alpha > 1$, let $k_n = [\alpha^n]$ be the closest integer to α^n, then for some constant c we have*

$$\sum_{n:k_n \geq i} \frac{1}{k_n^2} \leq \frac{c}{i^2}$$

uniform in i.

Proof. We cannot say it holds just because the sequence converge, since we want a constant uniform in i. The way to show the result is this:

$$\sum_{n:k_n \geq i} \frac{1}{k_n^2} \leq \frac{1}{i^2}\sum_{n=0}^{\infty} \beta^{-n} \leq \frac{c}{i^2}$$

where β is such that $\dfrac{k_{n+1}}{k_n} \geq \beta > 1$ for n large enough. This is nothing but a straight forward construction of the sum, where really β is selected to balance out $\dfrac{i^2}{\alpha^{2(n+s)}}$ terms. This is just a tail of the series so we can always do it in element methods anyway. □

Now we prove the strong law of large numbers.

Proof.

First, we assume X_i are non-negative since we can separate the positive and negative part of X_i and deal with each separately.

Step 1: Truncation

Define the truncation of X_i to be

$$Y_i := X_i \mathbb{1}_{X_i < i}.$$

Then, we have that

$$\sum_{i=1}^{\infty} \mathbb{P}(X_i \neq Y_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(X_i \geq i) = \sum_{i=1}^{\infty} \mathbb{P}(X_1 \geq i) \leq \mathbb{E}[X_1] < \infty$$

where the second inequality above is because

$$\begin{aligned}
\mathbb{E}[X_1] = \int_0^\infty x d\mu_X &= \sum_{i=0}^\infty \int_i^{i+1} x d\mu_X := \sum_{i=0}^\infty \int_{\omega : i \le X(\omega) < i+1} X(\omega) d\mu_X \\
&\ge \sum_{i=0}^\infty \int_{\omega : i \le X(\omega) < i+1} \min\{X(\omega)\} d\mu_X = \sum_{i=0}^\infty i \int_{\omega : 1 \le X(\omega) < i+1} 1 d\mu_X \\
&= \sum_{i=0}^\infty i \mathbb{P}(i \le X < i+1) = \sum_{i=1}^\infty i \mathbb{P}(i \le X_1 < i+1) = \sum_{i=1}^\infty \mathbb{P}(X_1 \ge i)
\end{aligned}$$

and the last is by assumption.

But then by Borel Cantelli we get that

$$\mathbb{P}(X_i \ne Y_i \ \ i.o.) = 0$$

so it's enough to prove

$$\frac{1}{n} \sum_{i=1}^n Y_i \overset{as}{\to} \mathbb{E}[X_i].$$

Step 2: Truncation's expectation:

$$\mathbb{E}[Y_i] - \mathbb{E}[X_i] = \mathbb{E}[X_1 \cdot \mathbb{1}_{x_1 \ge i}] \to 0$$

which is simply by $\mathbb{E}[X_i] < \infty$, in particular we have

$$\frac{1}{n} \mathbb{E}[Y_i] \to \lim_{n \to \infty} \mathbb{E}[Y_i] = \mathbb{E}[X_i]$$

since the limit exists. Thus, it is enough to show that

$$Z_n \overset{as}{\to} 0$$

where

$$Z_n = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \mathbb{E}[Y_i] \right)$$

since we did nothing but moving the term $\mathbb{E}[Y_i]$ to the left and passing into the limit.

Step 3: reduce to convergence of only a certain subsequence of Z_n:

The idea is to show for any $\alpha > 1$, let $k_n : [\alpha^n]$ be the closest integer to α^n, then we show

$$Z_{k_n} \overset{as}{\to} 0.$$

As for why this is sufficient we will leave to step 4.

The heuristic of this choice of subsequence is because we want to find a polynomial growth for which the tail does not matter, thus show convergence.

To show this, we first fix ε, then we use Borel Cantelli to show that the following sets $\mathbb{P}(|Z_{k_n}| \geq \varepsilon)$ has null limsup:

$$\mathbb{P}(|Z_{k_n}| \geq \varepsilon) = \mathbb{P}\left(\left|(Y_1 - \mathbb{E}[Y_1]) + \cdots + (Y_{k_n} - \mathbb{E}[Y_{k_n}])\right| \geq \varepsilon \cdot k_n\right)$$

$$\text{(Chebyshev)} \leq \frac{1}{\varepsilon^2 k_n^2} \operatorname{Var}\left(\sum_{i=1}^{k_n} Y_i\right) \overset{pair.ind.}{=} \frac{1}{\varepsilon^2 k_n^2} \sum_{i=1}^{k_n} \operatorname{Var}\left(Y_i\right)$$

and we sum on n to get

$$\sum_{n=1}^{\infty} \mathbb{P}(|Z_{k_n}| \geq \varepsilon) \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 k_n^2} \sum_{i=1}^{k_n} \operatorname{Var}\left(Y_i\right) \overset{rearrange}{=} \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \operatorname{Var}(Y_i) \sum_{n: k_n \geq i} \frac{1}{k_n^2}$$

where by Lemma 10.2 we can further get

$$\sum_{n=1}^{\infty} \mathbb{P}(|Z_{k_n}| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \frac{\operatorname{Var}(Y_i)}{i^2} \leq \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \frac{\mathbb{E}[Y_i^2]}{i^2}$$

so we try to bound the last sum.

Now, to bound the sum we first use integral identity (prop 9.9) to write

$$\mathbb{E}[Y_k^2] = \int_0^{\infty} pt^{p-1} \mathbb{P}(|Y_k^2| \geq t) dt = \int_0^{\infty} pt^{2p-2} \mathbb{P}(|Y_k^2| \geq t^2) dt^2$$

and plugging in $p = 1$ we get

$$\mathbb{E}[Y_k^2] = \int_0^{\infty} 1t^{2-2} \mathbb{P}(|Y_k^2| \geq t^2) 2t dt = 2\int_0^{\infty} t\mathbb{P}(|Y_k| > t) dt \leq 2\int_0^{k} t\mathbb{P}(|X_1| > t) dt$$

by definition of the truncation Y_k. And we write that upper integral bound with indicator function, and going back to the sum we get

$$\sum_{i=1}^{\infty} \frac{\mathbb{E}[Y_i^2]}{i^2} \leq \sum_{i=1}^{\infty} \frac{1}{i^2} \int_0^{\infty} \mathbb{1}_{t<i} \cdot 2t \cdot \mathbb{P}(|X_1| > t) dt = \int_0^{\infty} \left(\sum_{i=1}^{\infty} \frac{2t}{i^2} \cdot \mathbb{1}_{t<i} \cdot \mathbb{P}(|X_1| > t)\right) dt$$

where we can exchange limit because everything is positive and well-defined (or by Fubini). But notice that for $t > 0$

$$2t \sum_{i>t} \frac{1}{i^2} \leq 4$$

by computation, so we have

$$\sum_{i=1}^{\infty} \frac{\mathbb{E}[Y_i^2]}{i^2} \leq \int_0^{\infty} \left(\sum_{i=1}^{\infty} \frac{2t}{i^2} \cdot \mathbb{1}_{t<i} \cdot \mathbb{P}(|X_1| > t)\right) dt \leq \int_0^{\infty} 4\mathbb{P}(|X_1| > t) dt = 4\mathbb{E}[X_1] < \infty.$$

Note that this means, concluding all above, that the sets

$$\sum_{n=1}^{\infty} \mathbb{P}(|Z_{k_n}| \geq \varepsilon) < \infty$$

and by Borel Cantelli their limsup is 0. This means that the probability measure of such points for which it doesn't converge to 0 is 0, so $Z_{k_n} \xrightarrow{as} 0$.

Step 4: Show that $Z_n \xrightarrow{as} 0$:

The idea is to use sandwich, for which we define

$$T_n = \sum_{i=1}^{n} Y_i.$$

Then we first fix α, then fix some m such that $k_n \leq m \leq k_{n+1}$ for n large enough. We then have

$$\frac{1}{\alpha}\frac{T_{k_n}}{k_n} \leftarrow \frac{T_{k_n}}{k_n}\frac{k_n}{k_{n+1}} = \frac{T_{k_n}}{k_{n+1}} \leq \frac{T_m}{m} \leq \frac{T_{k_{n+1}}}{k_n} = \frac{T_{k_{n+1}}}{k_{n+1}}\frac{k_{n+1}}{k_{n+1}} \to \alpha\frac{T_{k_{n+1}}}{k_{n+1}}$$

and if we "let $\alpha \to 1$" we'll be done. To write things out formally we let

$$\mu := \lim_{n\to\infty} \frac{T_{k_n}}{k_n}$$

then we have

$$\frac{\mu}{\alpha} \leq \liminf_{m\to\infty} \frac{T_m}{m} \leq \limsup_{m\to\infty} \frac{T_m}{m} \leq \alpha\mu$$

for $\forall \alpha > 1$. Then we can actually let $\alpha \to 1$ and get that indeed

$$\lim_{n\to\infty} \frac{T_n}{n} = \lim_{n\to\infty} \frac{T_{k_n}}{k_n}$$

which really is what we need since

$$\frac{T_n}{n} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Hence $Z_n \xrightarrow{as} 0$ and we are done. □

10.1. **Characteristic functions.**

We need to deal with characteristic functions since (as will be shown in next class) Levy's continuity theorem says that $X_n \xrightarrow{d} X$ iff $\phi_{X_n} \to \phi_X$ pointwise. So we'd rather get a good understanding of characters first.

Def 10.3. *For a random variable X, the character is*

$$\phi_x(t) := \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tx) + i\sin(tx)].$$

Some properties of character functions are the following.

Proposition 10.4. *All character functions satisfy:*

(a) $\phi(0) = 1$.

(b) $\phi(-t) = \overline{\phi(t)}$ *due to* sin, cos *expression.*
(c) $|\phi(t)| \le \mathbb{E}\left[\left|e^{itX}\right|\right]$.
(d) $|\phi(t+h) - \phi(t)| \le \mathbb{E}\left[\left|e^{ihX} - 1\right|\right]$.
(e) $\mathbb{E}\left[e^{it(aX+b)}\right] = e^{itb}\phi_X(at)$.
(f) *For independent* X_1, X_2*,* $\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t)$.

Lemma 10.5. *For* $X \sim N(0,1)$*, we have* $\phi_X(t) = e^{-\frac{t^2}{2}}$.

Proof. We write it as an ODE form to get the solution.

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \int_{\mathbb{R}} \cos(tx) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

since sin is an odd function. Since $e^{-x^2/2}$ is in the integral it's asking us to do integral by parts, and we take the derivative to get

$$\frac{d}{dt}\left(\cos(tx)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\right) = -x\sin(tx)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

and hence by homework on derivatives we get

$$\frac{d}{dt}\phi_X(t) = \int_{\mathbb{R}} -x\sin(tx)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \sin(tx) d e^{-\frac{x^2}{2}}$$

with integral by part

$$= 0 - \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}} d\sin(tx) = -\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} t\cos(tx)e^{-\frac{x^2}{2}} dx = -t\phi_X(t).$$

where as we solve the ODE $\phi' = -t\phi$ by noticing

$$e^{t^2/2}\cdot\left(\phi' + t\phi\right) = \frac{d}{dt}\left(\phi e^{t^2/2}\right) = 0$$

which means that $\phi_X(t)e^{t^2/2}$ is a constant, thus

$$\phi_X(t) = ce^{-t^2/2}$$

and plugging in initial value we know $c = 1$ thus

$$\phi_X(t) = e^{-t^2/2}.$$

$\square$

11. 2/15: Inversion formula; Levy's continuity theorem

11.1. **Inversion formula.**

Theorem 11.1. *(Inversion formula) Let X be random variable with characteristic function ϕ_X. For $\theta > 0$ (just to write out a well defined integration), we define*

$$f_\theta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx-\theta t^2}\phi_X(t)dt$$

then for any bounded continuous $g : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}[g(X)] = \lim_{\theta\to 0}\int_{-\infty}^{\infty} g(x)f_\theta(x)dx.$$

Corollary 11.2. *Two random variables have the same characteristic function $\iff$ they have the same law.*

Proof.

($\Rightarrow$:) This direction is direct by theorem 11.1 since if they have the same character, then

$$\mathbb{E}[g(X)] = \mathbb{E}[g(Y)]$$

for all bounded continuous g, which is the equivalent definition of convergence in distribution, hence they have the same law (which is distribution).

($\Leftarrow$:) If they have the same law, then by definition of character they have the same integral, so it's the same. $\square$

Proof. (Theorem 11.1)

Let μ be the law of random variable X, then by definition

$$\phi_X(t) = \int_{\mathbb{R}} e^{ity}d\mu(y).$$

Fubini then give

$$f_\theta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it(y-x)-\theta t^2}dtd\mu(y)$$

where we can use Fubini because the integrand is on the unit ball (e^ik) and hence surely bounded. Then we do the change of variable $s = t\sqrt{2\theta}$ to get that the inner integral is

$$\frac{1}{\sqrt{2\theta}}\int_{-\infty}^{\infty} e^{i(y-x)\frac{s}{\sqrt{2\theta}}}e^{-\frac{s^2}{2}}ds$$

where we plug in the formula (character of normal gaussian)

$$\int_{\mathbb{R}} e^{itx}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx = e^{-\frac{t^2}{2}}$$

by letting $t = \dfrac{(y-x)}{\sqrt{2\theta}}$ and scaling by $\sqrt{2\pi}$ we get

$$f_\theta(x) = \frac{1}{\sqrt{2\theta}} \int_{-\infty}^{\infty} e^{i(y-x)\frac{s}{\sqrt{2\theta}}} e^{-\frac{s^2}{2}} ds = \sqrt{\frac{\pi}{\theta}} e^{-\frac{(y-x)^2}{4\theta}}.$$

Plugging in

$$f_\theta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sqrt{\frac{\pi}{\theta}} e^{-\frac{(y-x)^2}{4\theta}} d\mu(y) = \frac{1}{\sqrt{4\pi\theta}} \int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{4\theta}} d\mu(y).$$

Now recall result from homework 5, problem 1 that for X, Y independent where Y is continuous with density g, then $X+Y$ is continuous with density $\mathbb{E}[g(z-X)]$. Therefore f_θ is the density of $X + Y_\theta$ for $Y_\theta \sim N(0, 2\theta)$ as

$$\mathbb{E}[g(z_\theta - X)] = \frac{1}{\sqrt{4\pi\theta}} \int_{-\infty}^{\infty} e^{-\frac{(z_\theta - x)^2}{4\theta}} d\mu(x) = \frac{1}{\sqrt{4\pi\theta}} \int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{4\theta}} d\mu(y).$$

Now, for any continuous bounded g and $Z_\theta = X + Y_\theta$

$$\mathbb{E}[g(Z_\theta)] = \int_{-\infty}^{\infty} g(x) f_\theta(x) dx$$

and we see that our goal is to show $Z_\theta \xrightarrow{d} X$ as $\theta \to 0$ due to equivalent definition of convergence in distribution. But as $2\theta \to 0$ by chebyshev's inequality we get that

$$\mathbb{P}(|Y_\theta| > t) \leq \frac{\mathrm{Var}(Y)}{t^2} \to 0$$

and hence $Y_\theta \xrightarrow{p} 0$, hence $Z_\theta \xrightarrow{p} X$, which means $Z_\theta \xrightarrow{d} X$ and we are done. □

11.2. **Levy's continuity theorem.**

We will proceed as follows. We first introduce tightness and Helly's theorem, then prove Levy's theorem, and then fill up some holes (lemmas) in the proof.

Def 11.3. *A sequence of random variable is tight if for any $\varepsilon > 0$, $\exists k$ such that*

$$\sup_n \mathbb{P}(|X_n| \geq k) \leq \varepsilon.$$

Proposition 11.4. *(Helly's selection theorem) If $\{X_n\}$ is a tight family, then there exists a subsequence that converges in distribution.*

Proof.

Let F_n be the subsequence of X_n, then since F_n is a bounded function we can extract convergent subsequence n_k using compactness. Then we do the proof with the following four steps:

(1) Define
$$F_*(q) = \lim_{n_k \to \infty} F_{n_k}(q)$$
for all $q \in \mathbb{Q}$.

(2) Define
$$F_*(x) = \inf_{q \in \mathbb{Q}; q > x} F_*(q).$$

(3) Claim that F_* is the cdf of some random variable.

(4) Show $F_{n_k} \to F_*$ point wise at continuity points.

Now we do it step by step. To show (1) is fine we need to first construct sequence for each $q \in \mathbb{Q}$, then use diagonal argument to get a subsequence that work for all q.

For (2) we check that it is monotone, which also help us show that F_* is monotone everywhere. Also, it's right continuous by it's construction.

For (3) we already have monotone and right continuous, left limit exist, then we only need to check limits. But tightness means that $F(t) \to 1$ as $t \to \infty$ since
$$1 - F_n(k) + F_n(-k) = \mathbb{P}(|X_n| \geq k) < \varepsilon.$$
So we really finished checking (3).

For (4) it's just regular analysis.

□

Lemma 11.5. *Let X be a random variable with character ϕ_X, then for all $t > 0$ we have*
$$\mathbb{P}(|X| \geq t) \leq \frac{t}{2} \int_{-t/2}^{t/2} (1 - \phi_X(s)) ds.$$

Theorem 11.6. *(Levy's continuity theorem)*
$$X_n \xrightarrow{d} X \iff \phi_{X_n}(t) \to \phi_X(t) \quad \textit{point wise.}$$

Proof.

($\Rightarrow$:)

If $X_n \xrightarrow{d} X$, then convergence holds for continuous bounded function e^{itx}, so the point wise limit $\phi_{X_n}(t) \to \phi_X(t)$ holds by equivalent definition of conv in distribution.

($\Leftarrow$:)

Assume $\phi_{X_n}(t) \to \phi_X(t)$ point wise, then for fixed $\varepsilon > 0$ we can choose a small such that
$$\frac{1}{a} \int_{-a}^{a} (1 - \phi_X(s)) ds \leq \varepsilon$$
as long as $|\phi_X(s) - 1| \leq \frac{\varepsilon}{2}$. But the latter holds because $\phi_X(s)$ is continuous and is 1 at $s = 0$. Thus we really have the integral inequality above.

Now, by DCT

$$\lim_{n\to\infty} \frac{1}{a}\int_{-a}^{a}(1-\phi_{X_n}(s))ds = \frac{1}{a}\int_{-a}^{a}(1-\phi_X(s))ds \leq \varepsilon.$$

Then we can pick $t := \dfrac{2}{a}$ and by Lemma 11.5 we have

$$\limsup \mathbb{P}(|X_n| \geq t) \leq \varepsilon$$

for large t (really for small a). And thus getting rid of the lim sup we get

$$\mathbb{P}(|X_n| \geq t) \leq 2\varepsilon$$

and thus X_n is a tight family. Then by Prop 11.5 (Helly's) we get that there exists a converging subsequence of X_n in distribution. We claim that the limit of this subsequence is X, since if we assume the contrary and the limit is $Y \neq X$, we will get by the $\Rightarrow$ part of Levy's (just other direction of this theorem)

$$\phi_X(t) = \phi_{X_{\psi(n)}}(t) \to \phi_Y(t)$$

then by Corollary 11.2 we get that X and Y has the same law.

Now we try to construct a contradiction using above. Assume $X_n \xrightarrow{d} X$ does not hold, then it follows that $\exists$ a bounded continuous f such that $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ is not true. Then, for any subsequence n_j, we have

$$\left|\mathbb{E}[f(X_{n_j})] - \mathbb{E}[f(X)]\right| > \varepsilon.$$

But note that X_{n_j} is still a tight family, so there exists a subsequence of it that converges. Again, it converges to a random variable that has the same law as X. Contradiction! Hence $X_n \xrightarrow{d} X$.

$\square$

12. 2/20: CLT; Random vectors; Polish spaces

In practice, we might need to check whether a function is a character function, for which we use the following:

Theorem 12.1. *(Bochner's Theorem): Assume that a function $\phi(t)$ satisfies that for any sequence $\{t_n\}$, we define the matrix*

$$A := (a_{ij})_{i,j=1}^{k}, \quad a_{j,k} = \phi(t_j - t_k)$$

and assume that A has the following properties:

(1) A is Hermitian;
(2) A is positive semi definite, i.e. $\bar{v}^T Av \geq 0$;
(3) ϕ is continuous at the origin with $\phi(0) = 1$.

Then, ϕ is a characteristic function.

Moreover, note that the conditions are necessary and sufficient.

12.1. CLT.

We state the theorem, introduce lemmas we use, then prove the theorem with lemmas.

Theorem 12.2. *(CLT for iid sums) Let X_i be iid random variables with mean μ and* $\mathrm{Var}(X_i) = \sigma^2$. *Then*

$$\frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} (X_i - \mu) \xrightarrow{d} N(0, 1).$$

Lemma 12.3. *For any $x \in \mathbb{R}$, $k \geq 0$, we have*

$$\left| e^{ix} - \sum_{j=0}^{k} \frac{(ix)^j}{j!} \right| \leq \frac{|x|^{k+1}}{(k+1)!}.$$

This is just by Taylor's residue theorem and the fact that all derivatives are uniformly bounded by 1 (in the expansion we have those terms).

Corollary 12.4.

$$\left| e^{ix} - 1 - ix + \frac{x^2}{2} \right| \leq \min\left\{ x^2, \frac{|x|^3}{6} \right\}.$$

Proof. For $\leq \dfrac{|x|^3}{6}$ we just apply Lemma 12.3 and notice that last bits are first 3 terms in the Taylor expansion. For the other upper bound we note that

$$\left| e^{ix} - 1 - ix + \frac{x^2}{2} \right| \leq \left| e^{ix} - 1 - ix \right| + \left| \frac{x^2}{2} \right| \leq x^2$$

again by direct application of Lemma 12.3. □

Lemma 12.5. *Let $\{a_n\}$ and $\{b_n\}$ be complex sequences with $|a_i| \leq 1$, $|b_i| \leq 1$. Then*

$$\left|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\right| \leq \sum_{i=1}^{n} |a_i - b_i|.$$

Proof. We note that

$$\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i = \sum_{r=1}^{n} \left(a_1 \cdots a_{r-1} b_r \cdots b_n - a_1 \cdots a_r b_{r+1} \cdots b_n\right)$$

by telescoping sum. But then we can easily bound non-variant terms by 1 and the result follows by taking the absolute value and trig inequality. □

Proof.

Of course we use Levy's continuity theorem, which let us first calculate the character.

First, we replace $X_i \mapsto \dfrac{X_i - \mu}{\sigma} =: \tilde{X}_i$ and define $S_n = \dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^{n} \tilde{X}_i$. Then what we need to show is $\phi_{S_n} \to \phi_{N(0,1)}$.

To compute this, we have

$$\phi_{S_n}(t) \overset{ind}{=} \prod_{i=1}^{N} \phi_{\tilde{X}_i}\left(\frac{t}{\sqrt{N}}\right) \overset{iid}{=} \left[\phi_{\tilde{X}_i}\left(\frac{t}{\sqrt{N}}\right)\right]^N$$

and offhand we have

$$\phi_{N(0,1)} = e^{-\frac{t^2}{2}} = \left(1 - \frac{t^2}{2N}\right)^N.$$

To compare these, we use lemma 12.5 to get

$$\left|\phi_{S_n}(t) - \left(1 - \frac{t^2}{2N}\right)^N\right| \leq N \left|\phi_{\tilde{X}_i}\left(\frac{t}{\sqrt{N}}\right) - \left(1 - \frac{t^2}{2N}\right)\right|$$

if we let N large such that $t^2 \leq 2N$ and due to $\phi_{\tilde{X}_i}\left(\dfrac{t}{\sqrt{N}}\right) \leq 1$ (definition).

Hence we now use definition to re write the above as an expectation

$$\text{(previous)} = N \left|\mathbb{E}\left[e^{it\tilde{X}_i/\sqrt{N}} - 1 - \frac{it\tilde{X}_i}{\sqrt{N}} + \frac{t^2\tilde{X}_i^2}{2N}\right]\right|$$

where we can insert because $\mathbb{E}\left[\frac{it\tilde{X}_i}{\sqrt{N}}\right] = 0$ as $\mathbb{E}[\tilde{X}_i] = 0$ and we can reform the last term as $\mathrm{Var}(X_i) = 1$. Now we see we use corollary 12.4 to bound the above expectation by

$$\text{(previous)} \leq N\left|\min\left\{\frac{t^2\tilde{X}_i^2}{N}, \frac{|t|^3|\tilde{X}_i|^3}{6\sqrt{N^3}}\right\}\right| = \min\left\{t^2\tilde{X}_i^2, \frac{|t|^3|\tilde{X}_i|^3}{6\sqrt{N}}\right\}$$

and we note that since $\mathbb{E}[|\tilde{X}_i|^2] \leq \mathrm{Var}(\tilde{X}_i) < \infty$ we can use DCT to pass the limit. Once we've passed the limit we can then bound it using the second part of the upper bound and get that the difference of the character goes to 0 uniformly as $N \to \infty$. Then by Levy's we are done. $\square$

We now look at a most general version of the theorem, which uses the exact same way, module technical detail of bounding things, then see how it applies to a slightly weaker theorem.

Theorem 12.6. *(Linderberg-Feller CLT) Let $\{K_n\}$ be a sequence of positive integers that goes to ∞. For each n, let $\{X_{n,i}\}_{i=1}^{k_n}$ be the collection of iid random variables. Denote $\mu_{n,i} = \mathbb{E}[X_{n,i}]$, $\sigma_{n,i}^2 = \mathrm{Var}(X_{n,i})$, and $S_n^2 = \sum_{i=1}^{k_n}\sigma_{n,i}^2$. We can visualize this as a triangle of random variables with the n-th row having k_n random variables.*

Now we assume the Linderberg condition: suppose for $\forall\varepsilon > 0$ we have

$$\lim_{n\to\infty}\frac{1}{S_n^2}\sum_{i=1}^{k_n}\mathbb{E}\left[(X_{n,i} - \mu_{n,i})^2 \cdot \mathbb{1}_{|X_{n,i}-\mu_{n,i}|\geq\varepsilon\cdot S_n}\right] = 0.$$

Then

$$\frac{\sum_{i=1}^{k_n}(X_{n,i} - \mu_{n,i})}{S_n} \xrightarrow{d} N(0,1).$$

Corollary 12.7. *(Lyapunov's CLT) Let $\{X_n\}$ be the sequence of random variables with $\mu_i = \mathbb{E}[X_i]$ and $\sigma_i^2 = \mathrm{Var}(X_i)$. Let $S_N^2 := \sum_{i=1}^N \sigma_i^2$. Now, if for some $\delta > 0$ we have the Lyapunov condition:*

$$\lim_{n\to\infty}\frac{1}{S_n^{2+\delta}} \cdot \sum_{i=1}^{n}\mathbb{E}\left[|X_i - \mu_i|^{2+\delta}\right] = 0.$$

Then

$$\frac{\sum_{i=1}^{N}}{S_n} \xrightarrow{d} N(0,1).$$

It's left as an exercise to show that the Lyapunov condition implies the Linderberg condition.

12.2. **Random vectors:**

Def 12.8.

- *A random vector is a measurable map with*

$$f : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^n.$$

We can write it as $x = (x_1, \dots, x_n)$ for each x_i measurable.

- *Then we define it's cdf by*

$$F_x = (t_1, \dots, t_n) := \mathbb{P}(x_1 \le t_1, \dots, x_n \le t_n)$$

which we also call the joint cdf for $x_1, \dots, x_n$.

- *It's probability density function, if exists, is*

$$f_x : \mathbb{R}^d \to [0, \infty)$$

such that $\forall A \in \mathcal{B}(\mathbb{R}^n)$ we have

$$\mathbb{P}(x \in A) = \int_A f_x dt_1 \cdots dt_n$$

where the measure here is the Lebesgue measure $dt_1 \cdots dt_n$.

Def 12.9. *The (non-degenerate) multivariate Gaussian random vector given $\mu \in \mathbb{R}^d$ and strictly positive definite ($v^T \Sigma v \ge 0$) Σ is defined via the below function*

$$f_G := \frac{1}{(2\pi)^{1/2} \cdot (\det \Sigma)^{1/2}} \cdot \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right)$$

which is a density function. We call the corresponding random variable the multivariate Gaussian. We denote it by

$$X \sim N_d(\mu, \Sigma).$$

Proposition 12.10. *X is multivariate Gaussian if $a \cdot x$ is Gaussian for any $a \in \mathbb{R}^d$.*

We note that really it is a iff relationship if we were to exclude the degenerate case (where $x = 0$ for some component of x).

Recall that yet another equivalent definition is that

$$X = DW + \mu$$

where D is a matrix, W is the matrix of independent normals, and μ is the mean vector.

12.3. **Polish spaces.**

What we want to do now is to extend the definition of random variables to maps whose range is not $\mathbb{R}$ but more general metric spaces.

Def 12.11. *A metric space is a pair (M, d) with $d : M \times M \to \mathbb{R} \geq 0$ with*

(1) $d(x, y) = 0 \iff x = y$.
(2) $d(x, y) = d(y, x)$.
(3) $d(x, z) \leq d(x, y) + d(y, z)$.

Def 12.12. *Polish space is a separable completely metrizable topological space.*

Def 12.13. *M is completely metrizable if $\exists$ at least 1 choice of d such that (M, d) is a complete metric space.*

Def 12.14. *(M, d) is complete if Cauchy sequence converges.*

Def 12.15. *A sequence is Cauchy if $\forall x > 0$, $\exists N$ such that $d(x_m, x_n) < r$ if $m, n \geq N$.*

13. 2/22: Polish Space; Equivalent conditions of weak convergence; Brownian Motion

13.1. Polish space: random variables.

Recall that a polished space is a completely metrizable separable space.

Example 13.1.

(1) $\mathbb{R}^n$ is a polished space with Euclidean metric and countable dense subset $\mathbb{Q}^n$.
(2) $C[0,1]$ is a polished space with the infinite norm since first it can be approximated by polynomials with rational coefficients, which is a countable set; Moreover, we define it's pointwise limit and since uniform convergence passes continuity we get that the limit is in the space, then we show it is indeed the limit in the uniform norm with trig inequality to 3 parts. (recall proof that $C(K)$ is Banach.)

Def 13.2. *A random variable with values in polish space is any measurable function*

$$(\Omega, \mathcal{F}, \mathbb{P}) \mapsto (S, \rho)$$

where ρ is the metric.

Now note that the definition of convergence in probability, convergence a.s., and convergence in L^p is literally the same if we change the absolute value $|x_n - x| \mapsto \rho(x_n, x)$. But the cdf is not even defined so we use a more general definition, i.e. weak convergence.

Def 13.3. *Let (S, ρ) be a polished space, a sequence of S-valued random variable is said to converge weakly to S-valued random variable X if for all continuous bounded $f : S \to \mathbb{R}$ we have*

$$\lim_{n\to\infty} \mathbb{E}\left[f(x_n)\right] = \mathbb{E}\left[f(x)\right].$$

Remark 13.4.

Check the following. Check whether it's weak star or weak

This is related to functional analysis. Note first that the dual space of bounded measures are continuous functions. Now the weak converge above is really weak* convergence since probability convergence is not even a vector space (total measure is only 1). Thus, we can only extend it to the space of bounded measures, test weak* convergence with continuous bdd functions (bdd since space compact or other reasons) and conclude convergence. Here, the dual product is defined as

$$\langle f, \mu\rangle_{dual} = \int_S f \, d\mu.$$

Def 13.5. *(Push forward measure) the law of S valued random variable X is*

$$\mu_X(A) := \mathbb{P}(X \in A)$$

for all $A \in \mathcal{B}(S)$, for $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \rho)$.

Note that if you have $(S, \mathcal{B}(S))$, the one can only consider space of probability measure on S.

Def 13.6. *A sequence of probability measures μ_n on S converge weakly to μ on S if for any continuous bounded f we have*

$$\lim_{n\to\infty}\int_S f\,d\mu_n = \int_S f\,d\mu.$$

Theorem 13.7. *(Portmanteau Lemma) Let (S,ρ) be a polished space and $\{\mu_n\}$ be a sequence of probability measures on S. Then the following are equivalent:*

(1) $\mu_n \to \mu$ *weakly.*
(2) $\int_S f\,d\mu_n \to \int_S f\,d\mu$ *for f bounded and uniformly continuous.*
(3) $\int_S f\,d\mu_n \to \int_S f\,d\mu$ *for f bounded and Lipschitz continuous.*
(4) *For every closed set $F \subset S$, we have*

$$\limsup_{n\to\infty}\mu_n(F) \leq \mu(F).$$

(5) *For every open set $V \subset S$, we have*

$$\liminf_{n\to\infty}\mu_n(V) \geq \mu(V).$$

(6) *For every Borel $A \subset S$ such that $\mu(\partial A) = 0$, we have*

$$\lim_{n\to\infty}\mu_n(A) = \mu(A).$$

(7) $\int f\,d\mu_n \xrightarrow{n\to\infty} \int f\,d\mu$ *for every bounded measurable functions that is continuous a.e. with respect to μ.*

The proof is in Bi...textbook.

We only show (3) to (4) here.

Proof.

We fix $F \subset S$ and define

$$f_F(x) = \rho(X,F) := \inf_{y\in F}\rho(x,y).$$

Then we claim that f is Lipschitz: for $\forall x, x' \in S$ with $x \neq x'$ and $y \in F$ we have

$$\rho(x,y) \leq \rho(x,x') + \rho(x',y)$$

where we take infimum to get

$$f(x) \leq \rho(x,x') + f(x')$$

and for the reverse order we just use the other triangle inequality to get

$$f(x') \leq \rho(x,x') + f(x)$$

and then in total we get

$$\frac{|f(x) - f(x')|}{\rho(x,x')} \leq 1.$$

Now, since F is closed, we know that $f(x) = 0 \Rightarrow x \in F$. And our goal is to construct a sequence of bounded lipschitz continuous g_k for which we can pass our limit of the measures.

We construct g_k by

$$g_k(x) := (1 - kf(x))\Big|_+ := \max\{0, (1 - kf(x))\}$$

where since f is a non-negative function $g_k \in [0,1]$ and has Lipschitz constant k. Moreover, $g_k \downarrow \mathbb{1}_F$ pointwise and $\mathbb{1}_F \leq g_k$.

Now we have

$$\limsup_{n\to\infty} \mu_n(F) \leq \lim_{k\to\infty} \limsup_{n\to\infty} \int g_k d\mu_n \overset{(3)}{\to} \lim_{k\to\infty} \int g_k d\mu \overset{DCT}{=} \int \mathbb{1}_F d\mu = \mu(F).$$

holds for all k, and DCT is just bounded by $\mathbb{1}$. □

Corollary 13.8. *Let (S,ρ) be a polished space. If for μ and ν both probability measures we have*

$$\int f d\mu = \int f d\nu$$

for all bounded continuous f, then $\mu = \nu$.

Proof. For any closed set F we have

$$\mu, \mu, \mu \ldots \overset{w*}{\rightharpoonup} \nu$$

and similar for the other direction

$$\nu, \nu, \nu \ldots \overset{w*}{\rightharpoonup} \mu$$

and so by (4) for any closed $F \subset S$ we know $\limsup \mu(F) = \mu(F) \leq \nu(F)$ and $\nu(F) \leq \mu(F)$, hence $\mu(F) = \nu(F)$. But the sigma algebra generated for all closed F is the Borel set of S, thus μ and ν agrees on $\mathcal{B}(S)$, since closed sets form a π system. □

13.2. **Brownian motions: definitions.**

Def 13.9. *For fixed $0 \leq t_1 \leq \cdots \leq t_k \leq 1$, define*

$$\Pi_{t_1,\ldots,t_k} : C[0,1] \to \mathbb{R}^k$$

such that

$$\Pi_{t_1,\ldots,t_k}(f) = (f(t_1), \ldots, f(t_k))$$

to be the <u>finite dimensional projection maps.</u>

Note that $\Pi_{t_1,\dots,t_k}$ is continuous in the sup norm, thus it is measurable.

Def 13.10. *The finite dimensional distributions of a probability measure μ on $C[0,1]$ are the push-forward measures of the projection maps. That is, for*

$$(\Omega, \mathcal{F}, \mu) \xrightarrow{\Pi} (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$$

and

$$\mu_*(A) = \mu(\Pi^{-1}(A)).$$

Theorem 13.11. *(Donsker's theorem) Let $\{X_i\}$ be i.i.d. random variables with zero means and variance 1. Define*

$$B_n\left(\frac{i}{n}\right) := \frac{1}{\sqrt{n}} \sum_{j=1}^{i} X_j$$

in other words, the function $B_n(t)$ is only defined on points $\in \left\{\frac{1}{n}, \frac{2}{n}, \dots, 1\right\}$. But we can just linearly fill in the middle parts and make it a continuous function. Then as $n \to \infty$, we have

$$B_n(t) \overset{weak*}{\rightharpoonup} B(t)$$

where $B(t)$ is a $C[0,1]$ valued random variable with $B(0) = 0$ finite dimensional distributions

$$(B(t_1), \dots, B(t_n))$$

being multivariate normal. Moreover, we know

$$\mathrm{Cov}(B(t_i), B(t_j)) = \min\{t_i, t_j\}.$$

Def 13.12. *The B defined above is the Brownian motion on $[0,1]$.*

14. 2/27: Building towards Donsker's theorem

The goal today is to understand when and how to use finite distributions to study convergence of probability measures on $C[0,1]$.

First, we know that $C[0,1]$ is a polished space with $||\cdot||_\infty$, then for any $0 \le t_1 < t_2 < \cdots < t_k \le 1$ we have

$$\Pi_{t_1,\dots,t_k}(f) = (f(t_1),\dots,f(t_k))$$

where $\mu \mapsto \mu^*$ is the push-forward measure. That is, the measure generated for all projections with all collections of points.

Def 14.1. *The Class C_f collection of finite dimensional sets is defined as*

$$C_f := \left\{ x \subset C[0,1] \middle| x = \Pi^{-1}_{t_1,\dots,t_k} H, H \in \mathcal{B}(\mathbb{R}^k), \forall k, \forall \{t_1,\dots,t_k\} \subset [0,1] \right\}.$$

Lemma 14.2. *C_f is a separating class, i.e. if μ and ν are 2 probability measures that agrees on C_f, then they agree on $\mathcal{B}(C[0,1])$.*

Proof.

Step 1: C_f is a π system:

We only need to show that it is closed under intersections.

We first consider for fixed $\{t_1,\dots,t_k\}$, then if $x = \Pi^{-1}_{t_1,\dots,t_k} H_x$, $y = \Pi^{-1}_{t_1,\dots,t_k} H_y$, then $x \cap y = \Pi^{-1}_{t_1,\dots,t_k}(H_x \cap H_y)$ which is also inside C_f.

Now we prove inductively on different $t_1,\dots,t_n$, for which we can only consider when there's 2 t_1, t_2. That is, if $x = \Pi^{-1}_{t_1,t_2}$ and if $t_1 < s < t_2$ (just for WLOG purpose) then we can use

$$\psi(x,y,z) = (x,z)$$

to get

$$\Pi^{-1}_{t_1,t_2} H_x = \Pi^{-1}_{t_1,s,t_2}\left(\psi^{-1}(H_x)\right)$$

and in the same way we can expand this to get all cases.

Step 2: We show that C_f generates the Borel σ-algebra:

To get the Borel σ-algebra we only have to generate all the closed norm balls, for which we represent as

$$\overline{B(x,\varepsilon)} := \bigcap_{r \in \mathbb{Q}\cap[0,1]} [y : |y(r) - x(r)| \le \varepsilon]$$

where really the thing inside is $\Pi_r^{-1}([x(r)-\varepsilon, x(r)+\varepsilon]) \in C_f$ and since this is a countable intersection we are done. ☐

However, the bad news is that to test against C_f this is not enough to check that $\mu_n \rightharpoonup \mu$.

Example 14.3.

Let's see one example. Now note that we are on the bounded (probability) measures on $C[0,1]$, thus we first define $z_n \in C[0,1]$ with

$$z_n = \begin{cases} nx & x \in [0, 1/n] \\ -nx + 2 & x \in [1/n, 2/n] \\ 0 & x \geq 2/n \end{cases}$$

so we have $z_n \to 0$ point wise yet not in the uniform norm.

Now we define the delta measure $\delta_{z_n} : \mathcal{B}(C[0,1]) \to [0,1]$ on the σ-algebra of function spaces as

$$\delta_{z_n}(A) = \begin{cases} 1 & z_n \in A \\ 0 & z_n \notin A \end{cases}$$

and we can show that $\delta_{z_n} \rightharpoonup \delta_0$ does not hold.

To do this, we find a test function f that the dual product $\langle f, \delta_{z_n} \rangle \to \langle f, \delta_0 \rangle$ does not hold.

Let

$$f(x) := \left(1 - \frac{1}{\varepsilon} d\left(x, \overline{B_{\varepsilon/2}(0)}\right)\right)_+$$

where we note that $d\left(x, \overline{B_{\varepsilon/2}(0)}\right)$ is the distance between x and $\frac{\varepsilon}{2}$ ball of 0, that is

$$d\left(x, \overline{B_{\varepsilon/2}(0)}\right) = \left(||x|| - \frac{\varepsilon}{2}\right)_+$$

so we can pick ε small enough so that $\frac{1}{\varepsilon} d\left(x, \overline{B_{\varepsilon/2}(0)}\right)$ for any x with $||x|| \geq \frac{1}{2}$ we have $f(x) = 0$, for which $\varepsilon \leq \frac{1}{4}$ is enough.

But then let's compute the dual:

$$\langle f, \delta_{z_n} \rangle = \int_{C[0,1]} f(x) d\delta_{z_n} \leq \int \mathbb{1}_{\{||x|| \leq 1/2\}} d\delta_{z_n} = \delta_{z_n}(\{||x|| \leq 1/2\}) = 0$$

since $z_n \notin \{||x|| \leq 1/2\}$.

Where as since $0 \in \{||x|| \leq \varepsilon/2\}$ we get

$$\langle f, \delta_0 \rangle = \int_{C[0,1]} f(x) d\delta_0 \geq \int \mathbb{1}_{\{||x|| \leq \varepsilon/2\}} d\delta_0 = \delta_0(\{||x|| \leq \varepsilon/2\}) = 1$$

which means $\delta_{z_n} \rightharpoonup \delta_0$ does not hold.

Yet on the other hand, we want to show that the measures $\lim_{n\to\infty} \delta_{z_n}$ and δ_0 are indistinguishable over C_f. That is, we want to show that

$$\delta_{z_n}(S) \to \delta_0(S)$$

for all $S \in C_f$.

So for all $S \in C_f$, by definition $S = \Pi^{-1}_{t_1,\dots,t_k} H$ for $H \in \mathcal{B}(\mathbb{R}^k)$. But note that $z_n \to 0$ point wise, so at each point in the collection $\{t_1, \dots, t_k\}$ we have $z_n \to 0$ (this convergence here actually attains exact limit after finite steps!). Moreover, there's only finite points so the convergence on these k is not only uniform but also attains exactly 0 after finite steps.

We can write out $H = \prod_{i=1}^k [a_i, b_i]$. If $0 \notin [a_i, b_i]$ for some i, then we know that both 0 is not in the preimage, nor was z_n for n sufficiently large: if the corresponding $t_i > 0$ then for large n, $z_n(t_i) = 0$; if $t_i = 0$, $z_n(0) = 0$.

Hence $\lim\limits_{n\to\infty} \delta_{z_n}(S) = 0 = \delta_0(S)$ for S corresponding to above H.

Now, for H such that $0 \in [a_i, b_i]$ for all i, then we know $z_n \in S$ for large n and $0 \in S$ by by definition. Thus, $\lim\limits_{n\to\infty} \delta_{z_n}(S) = 1 = \delta_0(S)$ in this case.

Combined, we see that the measures $\lim_{n\to\infty} \delta_{z_n}$ and δ_0 are indistinguishable over C_f.

Remark 14.4. *The remark here is that even if the limit of a sequence of measure μ_n agrees with another measure ν on C_f, we still do not have weak convergence of $\mu_n \rightharpoonup \nu$.*

But we can conquer this issue with the help of tightness.

Theorem 14.5. *If $\{\mu_n\}$ is a tight family of probability measures on $C[0,1]$ whose limiting distributions converges to finite dimension distributions of some measure μ, then $\mu_n \rightharpoonup \mu$ weakly.*

First, what is tightness here?

Def 14.6. *A collection of μ_n on (S, ρ) is tight if $\forall \varepsilon > 0$, $\exists K$ compact such that*

$$\sup_n \mu_n(S \backslash K) < \varepsilon.$$

This gives us a divide and conquer strategy to a very small term and on a compact set, which is perfect.

Theorem 14.7. *Let $\{\mu_n\}$ be a tight family of push-forward measures on $C[0,1]$, and suppose their finite dimensional distribution converges weakly. Then $\exists \mu$ such that $\mu_n \rightharpoonup \mu$ and the finite dimensional distributions converge to the finite dimensional distribution of μ.*

Def 14.8. *The <u>module of continuity</u> of f is*

$$W_f(\delta) = \sup_{|s-t|<\delta} |f(s) - f(t)|.$$

Proposition 14.9. *Let $\{X_n\}$ be a sequence of $C[0,1]$ random variable, then the sequence is tight if the following holds:*

(1) $\forall \varepsilon > 0$, $\exists a > 0$, *we have*

$$\mathbb{P}\left(|X_n(0)| > a\right) \leq \varepsilon.$$

(2) $\forall \varepsilon > 0$ *and* $\eta > 0$, $\exists \delta > 0$ *with*

$$\mathbb{P}\left(W_{X_n}(\eta)\right) \le \varepsilon.$$

My understanding is that the first condition is implies uniform bounded with the second, and the second also implies "Lipschitz" with a varying factor η.

Def 14.10. *For X a $C[0,1]$ valued random variable, then we call $X(t)$ for $t \in [0,1]$ as the coordinate random variable.*

Def 14.11. *$F \subset C[0,1]$ is equicontinuous if $\forall \varepsilon > 0$, $\exists \delta$ such that $|f(s) - f(t)| \le \varepsilon$, $\forall f \in F$ and $|s - t| \le \delta$.*

Proof. (Proposition 14.9)

The proof uses Arzela-Ascoli, i.e. a closed set $F \subset C[0,1]$ is compact iff it's uniformly bounded and equicontinuous.

Just for time we skip the proof. □

15. 3/1: DONSKER'S THEOREM

Def 15.1. *A family of probability measures Π on (S, ρ) is* <u>*relatively compact*</u> *if any sequence in Π attains a weak convergent subsequence.*

Note that this is just weak convergence.

Theorem 15.2. *(Prohorov's theorem): Π is tight $\iff$ it is relatively compact.*

Let's focus on what this tells us with respect to what we're doing. The following is our setting from last time:

Proposition 15.3. *Let $\{\mu_n\}$ be a sequence of probability measures on $C[0,1]$, and suppose we know that their finite dimensional distributions converge, i.e. we know*

$$\mu_{n,*}^{t_1,\dots,t_j} \rightharpoonup \mu_*^{t_1,\dots,t_j}.$$

Now if we further assume that $\{\mu_n\}$ is relatively compact, i.e. there exists subsequence n_k such that

$$\mu_{n_k} \rightharpoonup \nu$$

then

$$\mu_{n_k,*}^{t_1,\dots,t_j} \rightharpoonup \nu_*^{t_1,\dots,t_j}.$$

Proof. We skip the essential part as an exercise of change of variable.

To show the result, we need to show for all test functions that the convergence of dual product (integral) holds. So the main structure is to check

$$\int f \, d\mu_{n_k,*}^{t_1,\dots,t_j} \overset{f \rightsquigarrow g}{=} \int g \, d\mu_{n_k} \rightharpoonup \int g \, d\nu \overset{g \rightsquigarrow f}{=} \int f \, d\nu_*^{t_1,\dots,t_j}$$

where the middle step is just due to weak convergence, and the other 2 are by the change of variable, which is an exercise (really?). $\square$

Proposition 15.4. *Mapping theorem: If $\mu_n \rightharpoonup \mu$, then if $h : (S, \rho) \to (S', \rho')$ is a measurable map, and if for $D(h) = \{x : h \text{ is discontinuous at } x\}$ we have $\mu(D(h)) = 0$, then*

$$\mu_{n,*}^h \to \mu_*^h$$

where μ_*^h is the push forward measure with respect to h.

Personally I think this is just a "homeomoric" map h gives us a change of measure to different polish spaces.

As an extension of Proposition 15.3, we have:

Proposition 15.5. *If $\{\mu_n\}$ is relatively compact, i.e. there exists subsequence n_k such that*

$$\mu_{n_k} \rightharpoonup \nu \quad \text{and} \quad \mu_{n,*}^{t_1,\dots,t_j} \rightharpoonup \mu_*^{t_1,\dots,t_j}$$

then $\mu = \nu$ on $\mathcal{B}(C[0,1])$.

Proof.

For the first result, we use proposition 15.3 to get that

$$\mu_{n_k,*}^{t_1,\dots,t_j} \rightharpoonup \nu_*^{t_1,\dots,t_j}$$

and use Lemma below to get that for any n_k

$$\mu_{n_k,*}^{t_1,\dots,t_j} \rightharpoonup \mu_*^{t_1,\dots,t_j}$$

so being the limit of the same sequence we have

$$\mu_*^{t_1,\dots,t_j} = \nu_*^{t_1,\dots,t_j}.$$

For the last part, since C_f is a separating set, we get that $\mu = \nu$ on $\mathcal{B}(C[0,1])$ given that the first statement holds. □

Lemma 15.6. *We have $\mu_n \rightharpoonup \mu \iff$ each subsequence has a further subsequence convergent to μ.*

The Lemma means that it is enough to check weak convergence on the level of finite dimensional distributions for tight families.

Proof.

($\Rightarrow$:) This is just by definition of convergence.

($\Leftarrow$:) Assume that $\mu_n \not\rightharpoonup \mu$, then $\exists f$ bounded continuous such that

$$\int f\,d\mu_n = \langle f, \mu_n\rangle \not\rightharpoonup \langle f, \mu\rangle = \int f\,d\mu$$

which means that there is a subsequence of $\langle f, \mu_n\rangle$ that is far from $\langle f, \mu\rangle$. Now since f is bounded we know $\langle f, \mu_n\rangle$ lies within a bounded set, hence contained in a compact set in $\mathbb{R}$. So there is a subsequence that converges to something, which is surely not $\langle f, \mu\rangle$, contradiction. □

Theorem 15.7. *If $\{\mu_n\}$ is tight, assume that the finite dimensional distribution converges to some limit, then there exists a measure μ such that $\mu_n \rightharpoonup \mu$ and the corresponding limits are finite dimensional distributions for μ.*

Note that the most important thing is we can get a μ from the distributions.

Proof. For $\{\mu_n\}$ tight, there exists a subsequence and μ with $\mu_{n_k} \rightharpoonup \mu$. But this means

$$\mu_{n_k,*}^{t_1,\dots,t_j} \rightharpoonup \mu_*^{t_1,\dots,t_j}$$

moreover, by assumption we know

$$\mu_{n_k,*}^{t_1,\dots,t_j} \rightharpoonup l^{t_1,\dots,t_j}$$

which means

$$l^{t_1,\dots,t_j} = \mu_*^{t_1,\dots,t_j}.$$

But this is independent of n_k and since this works for all collection $\{t_1, \dots, t_j\}$ we know that μ has finite dimensional distributions that are exactly the limits of μ_n's, so we are done. $\square$

Now we restate Donsker's theorem, explain to an extent how this will be proven.

"Pollen particles written in water."

Theorem 15.8. *(Donsker's Theorem): Let $\{X_i\}$ be iid random variables with mean 0 and variance 1. So we can construct a sequence B_n that is a $C[0,1]$ valued random variable defined by*

$$B_n(t) = \begin{cases} 0 & t = 0 \\ \sum_{j=1}^{i} \frac{X_j}{\sqrt{n}} & t = \frac{i}{n} \\ \textit{linear interpolation} & \textit{otherwise}. \end{cases}$$

Then we know $B_n \rightharpoonup B$ where B is given in terms of its finite dimensional distribution, that is

- $B(0) = 0$*;*
- $[B(t_1), \dots, B(t_k)]$ *is a multivariate Gaussian with mean* 0 *and* $\mathrm{Cov}(B(t_i), B(t_j)) = \min\{t_i, t_j\}$.

Remark 15.9.

Note that what we mean by $B_n \rightharpoonup B$ really is $\mu_{B_n} \rightharpoonup \mu_B$, which is well-defined. Moreover, $\mu_{B,*}^{t_1,\dots,t_k}$ is the law of the multivariate Gaussian of the prescribed mean and covariance.

Def 15.10. *We call the above defined B Brownian motion and μ_B the Wiener measure.*

But let's see the standard definition of Wiener measure and see how this pops out as equivalent.

Def 15.11. *Let X be a $C[0,1]$ valued random variable, then we define $X(t)$ or X_t as the value of $X(\omega)$ at t. That is, $X(t)$ is a random variable $X_t(\omega)$ with value in $\mathbb{R}$.*

Def 15.12. *A Stochastic process is a collection of random variables $\{X_t\}_t$, where $t \in I$ the index set, which is usually $[0,1]$ and denotes time.*

Def 15.13. *The Wiener measure W is a probability measure on $C[0,1]$ satisfying the following:*

(1) $W[X(t) \le \alpha] = \dfrac{1}{\sqrt{2\pi t}} \displaystyle\int_{-\infty}^{\alpha} e^{-\frac{u^2}{2t}}\, du$ *for* $t \neq 0$. *For* $t = 0$ *we have* $W[X_0 = 0] = 1$.

(2) For $0 \le t_1 < \cdots < t_k \le 1$, *we know that* $X_{t_k} - X_{t_{k-1}}, X_{t_{k-1}} - X_{t_{k-2}}, \dots$ *are independent random variables.*

(3) (Stationary increment) $X(t) - X(s) \sim N(0, t-s)$ for $t \geq s$.

Note that the first condition is really an instance of the third case. But it is really important we see the exact expression there.

Proposition 15.14. *For X defined above, we know $(X(t_1), \dots, X(t_n))$ has the covariance described in theorem, i.e.* $\mathrm{Cov}(X(t_i), X(t_j)) = \min\{t_i, t_j\}$.

Proof. First, we show that they are independent. This is due to the decomposition

$$\begin{pmatrix} X(t_1) \\ X(t_2) \\ \vdots \\ X(t_{k-1}) \\ X(t_k) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 1 & 1 & 1 & \ddots & 0 \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} X(t_1) \\ X(t_2) - X(t_1) \\ \vdots \\ X(t_{k-1}) - X(t_{k-2}) \\ X(t_k) - X(t_{k-1}) \end{pmatrix}$$

and since the right most vector has independent elements, so does the vector on left, since linear map does preserves that. Now we can compute the covariance. Assume $t \geq s$, note that $\mathbb{E}[X(t)] = \mathbb{E}[X(s)] = 0$ we have

$$\begin{aligned} \mathrm{Cov}(X(t), X(s)) &= \mathbb{E}[X(t)X(s)] - \mathbb{E}[X(t)]\mathbb{E}[X(s)] = \mathbb{E}[X(t)X(s)] \\ &= \mathbb{E}[X^2(s) + X(s)(X(t) - X(s))] \\ &= \mathbb{E}[X^2(s)] + \mathbb{E}[X(s) - X(0)]\mathbb{E}[X(t) - X(s)] \\ &= s + 0 = s = \min\{s, t\} \end{aligned}$$

where the last line is because $\mathbb{E}[X^2(s)] = \mathrm{Var}(X(s) - X(0)) - \mathbb{E}[X(s)] = s$. ☐

Now we give the strategy of proving the Donsker's theorem:

(1) We first check that μ_n^B is a tight family. To show this we use Ascholi Arzela.
(2) We know that convergence of finite dimensional distribution $\iff$ convergence of random vectors. So we show the former with a vector version of CLT.

The second step contains some more steps in the generated CLT part, for which we look at just one results there.

Corollary 15.15. *(Corollary of multivariate Levy's theorem) (Cramer-Wold)*

Let $\{X_n\}$ be a sequence of m dimensional vector, then

$$X_n \rightharpoonup X \iff tX_n \xrightarrow{d} tX$$

for any test vector t.

Appendix A. a

Appendix B. b

Appendix C. c

Acknowledgements.