8. Weak convergence in $\mathbb{R}$

Weak convergence of measures on the real line is of course of basic interest — it appears in the central limit theorem. This lecture is not going to repeat what you know or will learn during courses in probability. We only mention a strong analytic technique of analyzing weak convergence of measures.

For a given $\mu \in P(\mathbb{R})$, its **characteristic function** $\widehat{\mu}$ is a function $\mathbb{R} \to \mathbb{C}$ defined as

$$\widehat{\mu}(t) = \int_{\mathbb{R}} e^{itx} \, \mathrm{d}\mu(x).$$

Here $e^{itx} = \cos(tx) + i\sin(tx) \in \mathbb{C}$. Outside probability, $\widehat{\mu}$ is rather called the Fourier transform of μ (with minus in the exponent).

Seeing this for the first time one can ask: 'Why I am supposed to integrate complex-valued functions?'. There are two reasons; the first is that this is actually not complicated — one can always integrate separately the real and the imaginary part. The complex form, however, is much more convenient for analytic manipulations. The second reason is that $\widehat{\mu}$ carries all the information about the measure μ. Once we know $\widehat{\mu}(n)$ for integer values of n then we know all the integrals $\int_{\mathbb{R}} \cos(nx) \, \mathrm{d}\mu(x)$ and $\int_{\mathbb{R}} \sin(nx) \, \mathrm{d}\mu(x)$. Consequently, we *know* the measure μ, that is we know $\int_{\mathbb{R}} g(x) \, \mathrm{d}\mu(x)$ for every $g \in C_b(\mathbb{R})$; see the problem list.

Note first that $\widehat{\mu}$ is a continuous function for every $\mu \in P(\mathbb{R})$, for if $t_k \to t$ then, by the Lebesgue dominated convergence theorem, $\widehat{\mu}(t_k) \to \widehat{\mu}(t)$.

Lemma 8.1. *Let $g = \widehat{\mu}$ for some $\mu \in P(\mathbb{R})$. Then for every $u > 0$ we have*

$$\frac{1}{u} \int_{-u}^{u} (1 - g(t) \, \mathrm{d}t \geq \mu\left(\{x : |x| \geq 2/u\}\right).$$

Proof. By the definition of a characteristic function (and obvious changes),

$$\frac{1}{u} \int_{-u}^{u} (1 - g(t)) \, \mathrm{d}t = 2 - \frac{1}{u} \int_{-u}^{u} \left(\int_{\mathbb{R}} e^{itx} \, \mathrm{d}\mu(x) \right) \, \mathrm{d}t =$$

changing the iterated integrals via Fubini,

$$= 2 - \frac{1}{u} \int_{\mathbb{R}} \left(\int_{-u}^{u} e^{itx} \, \mathrm{d}t \right) \, \mathrm{d}\mu(x) =$$

we calculate the inner inegral $1/u \int_{-u}^{u} e^{itx} \, \mathrm{d}t = 2\sin(ux)/(ux)$; now it is clear that everything is real; we get

$$= 2 - \int_{\mathbb{R}} 2 \frac{\sin ux}{ux} \mathrm{d}\mu(x) = 2 \int_{\mathbb{R}} \left(1 - \frac{\sin ux}{ux} \right) \mathrm{d}\mu(x) \geq$$

and, since we now integrate a nonnegative function,

$$\geq 2 \int_{\{x : |ux| \geq 2\}} \left(1 - \frac{\sin ux}{ux} \right) \mathrm{d}\mu(x) \geq 2 \cdot \frac{1}{2} \cdot \mu(\{x : |x| \geq 2/u\}),$$

finito! ☐

Lemma 8.2. *Let $g_n = \widehat{\mu_n}$ for some $\mu_n \in P(\mathbb{R})$. If g_n converge pointwise to a function g that is continuous at 0 then the family $\{\mu_n : n \in \mathbb{N}\}$ is uniformly tight.*

Proof. Fix $\varepsilon > 0$. Note that $g_n(0) = 1$ (by the definition of a characteristic function) for every n so $g(0) = 1$. As g is continuous at 0, there is $u > 0$ such that $|g(x) - 1| \le \varepsilon/4$ for $|x| < u$. Then

$$\frac{1}{u}\int_{-u}^{u} |1 - g(t)| \, \mathrm{d}t \le (1/u)\cdot(2u)\cdot\varepsilon/4 = \varepsilon/2.$$

Using the Lebesgue dominated convergence theorem we have

$$\frac{1}{u}\int_{-u}^{u} |1 - g_n(t)| \, \mathrm{d}t \le \varepsilon,$$

for large n, say that it happens whenever $n > n_0$.

Now by Lemma 8.1, we have

$$\mu_n([-2/u, 2/u]) \ge 1 - \varepsilon \text{ for } n > n_0.$$

Note that there is $\theta > 2/u$ such that $\mu_n[-\theta, \theta] \ge 1-\varepsilon$ for $n = 1, 2, \ldots, n_0$. Then $\mu_n[-\theta, \theta] \ge 1 - \varepsilon$ for every n, and this is what we wanted. ☐

Theorem 8.3. *For $\mu_n, \mu \in P(\mathbb{R})$ TFAE*

(i) $\mu_n \longrightarrow \mu$;

(ii) $\lim_n \widehat{\mu_n}(t) = \widehat{\mu}(t)$ *for every* $t \in \mathbb{R}$.

Proof. $(i) \to (ii)$ follows from the very definition of weak convergence, since the functions $\sin, \cos$ are bounded and continuous.

$(ii) \to (i)$ follows from Lemma 8.2 and Prokhorov's theorem: the set $\{\mu_n : n \in \mathbb{N}\}$ is relatively compact so every subsequence has a further subsequence that is converging in $P(\mathbb{R})$. But for every cluster point ν of that set of measures we have $\widehat{\nu} = \widehat{\mu}$ which means $\nu = \mu$, so the whole sequence must converge to μ. ☐

Those characteristic functions can be defined and successfully used for measures on Euclidean spaces: for $\mu \in P(\mathbb{R}^d)$ we define $\widehat{\mu} : \mathbb{R}^d \to \mathbb{C}$ by

$$\widehat{\mu}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} \, \mathrm{d}\mu(x),$$

where $\langle t, x\rangle = \sum_{k=1}^{d} t_k x_k$ is the scalar products.

9. There is (essentially) one measure — Intro on Boolean algebras

At this point one might quote Monty Python:
And now for something completely different [1]

After seeing whole spaces of measures it is good to realize that all nonatomic probability measures we discussed so far are incarnations of the Lebesgue measure on $[0, 1]$. The fact

[1] `https://www.youtube.com/watch?v=AB1pT1q1GqI&t=1919s` in case the quotation foes not ring a bell.

that the Haar measure on $2^{\mathbb{N}}$ can be transferred to λ was mention in L2/P10. The following holds for measures on nice spaces.

Theorem 9.1. *If X, Y are Polish spaces and $\mu \in P(X)$, $\nu \in P(Y)$ are nonatomic then there is a Borel isomorphism $f : X \to Y$ such that $f[\mu] = \nu$ (and $\mu = f^{-1}[\nu]$).*

This will be partially discussed on the next problem list. We shall outline below a more general phenomenon related to Boolean algebras.

Let us recall (or introduce) the concept of a Boolean algebra $\mathfrak{A}$, or more formally, $\langle \mathfrak{A}, \vee, \wedge, ^c 0, 1\rangle$. Here $\mathfrak{A}$ is a set containing two distinct elements $0, 1$, equipped with two binary operations $\vee, \wedge$, and an unary operation ^{c} with the intention that those behave exactly as the usual set-theoretic operation $\cup, \cap$ applied to subsets of some space X; ^{c} corresponds to the complement in X, $0 = \emptyset$, $1 = X$. It is not important to examine a really long list of axioms of a Boolean algebra — it is enough to understand that those axioms guarantee that every true formula, such as

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \quad \text{or} \quad A \cup (X \setminus A) = X,$$

has its Boolean equivalent; here

$$(a \vee b) \wedge c = (a \wedge c) \vee (b \wedge c) \quad \text{or} \quad a \vee a^c = 1.$$

Note, however, that the above refers only to **finite** operations! In a Boolean algebra $\mathfrak{A}$ we can define a partial order $\le$ mimicking inclusion:

$$a \le b \iff a \vee b = b \iff a \wedge b = a.$$

Given any family of sets $\mathcal{A}$ we can form the set $S = \bigcup \mathcal{A}$ (at least if we assume the usual axioms of set theory). Note that the union S can be defined as the smallest set containing every $A \in \mathcal{A}$. In a Boolean algebra $\mathfrak{A}$, given any family $B \subseteq \mathfrak{A}$, we say that $s = \bigvee B$ or $s = \sup B$ if $s \in \mathfrak{A}$ is the least upper bound of B (i.e. $b \le s$ for $b \in B$ and whenever $b \le s'$ for every $b \in B$ then $s \le s'$. If every nonempty[2] $B \subseteq \mathfrak{A}$ has the least upper bound then the algebra $\mathfrak{A}$ is said to be *complete*. In a complete algebra every nonempty $B \subseteq \mathfrak{A}$ has the greatest lower bound (defined in a similar manner), and this follows easily once we realize that $a \le b$ is equivalent to $b^c \le a^c$.

Example 9.2. A trivial example is any algebra $\mathcal{A}$ of subsets of some X (an algebra of sets is by definition closed under all finite set-theoretic operations). We can take $\mathcal{A} = \{\emptyset, X\}$ or $\mathcal{P}(X)$ in the extremal cases. As it was mentioned, $\emptyset$ plays the role of 0 and X is 1. Note that $\mathcal{P}(X)$ forms a complete Boolean algebra but if, for instance, $\mathcal{A}$ is the algebra of finite or co-finite subsets of $\mathbb{R}$ then $\mathcal{A}$ is not complete; for exmple, the family $\{\{x\} : x \in [0, 1]\}$ does not have the least upper bound in $\mathcal{A}$.

Example 9.3. The most important examples of Boolean algebras are defined in the following way. Take any algebra of sets $\mathcal{A} \subseteq \mathcal{P}(X)$ and choose some ideal $\mathcal{I} \subseteq \mathcal{A}$. Here by an ideal we mean a family such that $\emptyset \in \mathcal{I}, X \notin \mathcal{I}$, if $A, B \in \mathcal{I}$ then $A \cup B \in \mathcal{I}$ and if $A \in \mathcal{I}$ then $B \in \mathcal{I}$ for every $B \in \mathcal{A}$ such that $B \subseteq A$.

[2] for $B = \emptyset$ we seem to have $\sup B = 0$, $\inf B = 1$

Then we define the quotient algebra $\mathfrak{A} = \mathcal{A}/\mathcal{I}$ of equivalence classes $A/\mathcal{I}$ of the equivalence relation

$$A \sim B \iff A \triangle B \in \mathcal{I}.$$

This means that we identify the sets from $\mathcal{A}$ that differ by a set from the fixed ideal. We define the operations in a natural way:

$$A/\mathcal{I} \vee B/\mathcal{I} = (A \cup B)/\mathcal{I}, \quad A/\mathcal{I} \wedge B/\mathcal{I} = (A \cap B)/\mathcal{I}, \quad (A/\mathcal{I})^c = (X \setminus A)/\mathcal{I},$$

and so on. Please check on some examples that those operations are well-defined and satisfy typical axioms.

The simplest example of that type: take the whole of $\mathcal{P}(\mathbb{N})$ and divide it by the ideal $\mathcal{I}$ of all finite subsets of $\mathbb{N}$. To see that such a Boolean algebra is quite tricky see the problem list.

Example 9.4. Finally, our main hero: take any probability measure space (X, Σ, μ) and $\mathcal{N} = \{A \in \Sigma : \mu(A) = 0\}$. Note that $\mathcal{N}$ is an ideal in th eabove sense; actually $\mathcal{N}$ is a σ-ideal, i.e. it is closed under countable unions.

The Boolean algebra $\mathfrak{A} = \Sigma/\mathcal{N}$ is called **the measure algebra** of μ. Here we finally do what we always wanted: To ignore the sets of measure zero. This object is in fact quite familiar; note that $\mathfrak{A}$ may be seen as the family of $\{0, 1\}$-'functions' from $L_1(\mu)$.

The measure algebra of some measure has a number of additional structures; we mention below the first one.

Theorem 9.5. *Let $\mathfrak{A} = \Sigma/\mathcal{N}$ be the measure algebra built from some probability measure space (X, Σ, μ) where $\mathcal{N}$ is the σ-ideal of sets of measure zero.*

(a) The algebra $\mathfrak{A}$ is σ-complete, that is every countable set $B \subseteq \mathfrak{A}$ has the least upper bound.
(b) The formula $\mu'(A/\mathcal{N}) = \mu(A)$ defines a function $\mu' : \mathfrak{A} \to [0, 1]$.
(c) μ' is countably additive in this sense, that if $a_n \in \mathfrak{A}$ and $a_n \wedge a_k = 0$ for $n \neq k$ then

$$\mu'\left(\bigvee_{n=1}^{\infty} a_n\right) = \sum_{n=1}^{\infty} \mu'(a_n).$$

(d) the measure μ' satisfies the equivalence $\mu'(a) = 0$ if and only if $a = 0$.
(e) The algebra $\mathfrak{A}$ is complete.

Proof. It is routine to check that if $A_n \in \Sigma$ and $a_n = A_n/\mathcal{N}$ then $\bigvee_n a_n$ is simply $\bigcup_n A_n/\mathcal{N}$.

To check (b), note that if $A_1/\mathcal{N} = A_2/\mathcal{N}$ then $A_1 \triangle A_2 \in \mathcal{N}$ which means that $\mu(A_1 \triangle A_2) = 0$ so $\mu(A_1) = \mu(A_2)$. This shows that μ' is well-defined.

(c) and (d) are easy exercises.

Perhaps (e) is a bit suprising; in Σ we may not be able to form an uncountable union. However, in the measure algebra we can. Take any (possibly uncountable) $B \subseteq \mathfrak{A}$. Using (a) we can define

$$r = \sup\{\mu'(\bigvee B_0) : B_0 \subseteq B, \text{ and } B_0 \text{ countable}\}.$$

Note that this supremum is attained, there is a countable $B_0 \subseteq B$ such that $r = \mu'(\bigvee B_0)$. It remains to check that $x = \bigvee B_0$ is the least upper bound for the whole family B. But for $a \in B$ we must have $a \leq x$ because, otherwise $a \setminus x \neq 0$ which implies (by (d)) $\mu'(a \setminus x) > 0$ and

$$\mu'(a \vee x) = \mu'(a \setminus x) + \mu'(x) > \mu(x) = r,$$

a contradiction. □