Random matrices and communication systems: WEEK 9

In this lecture, we first give a quick reminder regarding distributions on the real line; we then recall the notion of weak convergence of sequences of distributions and finally move to various characterizations of this weak convergence, more particularly in terms of moments and Stieltjes transform.

1 Distributions without random variables

1.1 Distributions on the real line

Let $\mathcal{B}(\mathbb{R})$ be the *Borel σ-field on* $\mathbb{R}$, that is, the smallest σ-field on $\mathbb{R}$ that contains all the open sets in $\mathbb{R}$; its elements $B \in \mathcal{B}(\mathbb{R})$ are called the *Borel sets*.[1] Recall that a mapping $f : \mathbb{R} \to \mathbb{R}$ is said to be *Borel-measurable* if for all $B \in \mathcal{B}(\mathbb{R})$, $f^{-1}(B) = \{x \in \mathbb{R} \ : \ f(x) \in B\} \in \mathcal{B}(\mathbb{R})$. In particular, any continuous function is Borel-measurable.[2]

Definition 1.1. A (probability) distribution on $\mathbb{R}$ is a mapping $\mu : \mathcal{B}(\mathbb{R}) \to [0,1]$ such that

$$\mu(\emptyset) = 0, \quad \mu(\mathbb{R}) = 1 \quad \text{and} \quad \text{if } (B_n,\, n \geq 1) \in \mathcal{B}(\mathbb{R}) \text{ are disjoint, then } \mu\left(\bigcup_{n \geq 1} B_n\right) = \sum_{n \geq 1} \mu(B_n)$$

Definition 1.2. The cumulative distribution function (cdf) associated to a distribution μ is the mapping $F_\mu : \mathbb{R} \to [0,1]$ defined as

$$F_\mu(t) = \mu((-\infty, t]), \quad t \in \mathbb{R}$$

Fact. The knowledge of the cdf F_μ is equivalent to that of the distribution μ.

There are two well known particular classes of distributions.

Discrete distributions, for which there exists a countable set C such that $\mu(C) = 1$. In this case,

$$\mu(B) = \sum_{x \in B \cap C} \mu(\{x\}) \quad \forall B \in \mathcal{B}(\mathbb{R})$$

and F_μ is a step function.

Continuous distributions, for which there exists a probability density fiunction (pdf) $p_\mu : \mathbb{R} \to \mathbb{R}_+$ such that

$$\mu(B) = \int_B p_\mu(x)\, dx \quad \forall B \in \mathcal{B}(\mathbb{R})$$

In this case, F_μ is a continuous function.

1.2 Lebesgue's integral

The Lebesgue integral of a Borel-measurable function $f : \mathbb{R} \to \mathbb{R}$ with respect to a distribution μ is defined in three steps as follows.

1. First suppose f is of the form

$$f(x) = \sum_{j \geq 1} y_j\, 1_{B_j}(x), \quad \text{where } y_j \geq 0 \text{ and } B_j \in \mathcal{B}(\mathbb{R}) \tag{1}$$

Then the integral is defined as

$$\int_{\mathbb{R}} f(x)\, d\mu(x) = \sum_{j \geq 1} y_j\, \mu(B_j)$$

[1] If one is not familiar with this notion, one may think of $\mathcal{B}(\mathbb{R})$ as being simply the set of (nearly!) all subsets of $\mathbb{R}$.
[2] Again, one may simply consider that (nearly!) all functions are Borel-measurable.

2. Next, suppose that f is Borel-measurable and non-negative (i.e. $f(x) \geq 0$ for all $x \in \mathbb{R}$). Define then for $n \geq 1$

$$f_n(x) = \sum_{j \geq 1} \frac{j-1}{2^n} \, 1_{\left\{x \in \mathbb{R} \,:\, \frac{j-1}{2^n} \leq f(x) < \frac{j}{2^n}\right\}}(x), \quad x \in \mathbb{R}$$

Then one can check that for all $n \geq 1$ and $x \in \mathbb{R}$, $f_n(x) \leq f_{n+1}(x)$ as well as $\lim_{n\to\infty} f_n(x) = f(x)$. As f_n is of the form (1), we may define

$$\int_{\mathbb{R}} f(x)\, d\mu(x) = \lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\, d\mu(x) = \lim_{n\to\infty} \sum_{j\geq 1} \frac{j-1}{2^n} \, \mu\left(\left\{x \in \mathbb{R} \,:\, \tfrac{j-1}{2^n} \leq f(x) < \tfrac{j}{2^n}\right\}\right)$$

Notice that as $f_n \leq f_{n+1}$, the above sequence is increasing, so the limit always exists, but may take the value $+\infty$.

3. Assume now that f is any Borel-measurable function. In this case case, we say that the integral is well defined only if

$$\int_{\mathbb{R}} |f(x)|\, d\mu(x) < \infty \tag{2}$$

and we set

$$\int_{\mathbb{R}} f(x)\, d\mu(x) = \int_{\mathbb{R}} f^+(x)\, d\mu(x) - \int_{\mathbb{R}} f^-(x)\, d\mu(x)$$

where $f^+(x) = \max(f(x), 0)$ and $f^-(x) = \max(-f(x), 0)$.

It is worth noticing that condition (2) is satisfied for any distribution μ when f is Borel-measurable and *bounded*, as

$$\int_{\mathbb{R}} |f(x)|\, d\mu(x) \leq \sup_{x\in\mathbb{R}} |f(x)| \int_{\mathbb{R}} d\mu(x) = \sup_{x\in\mathbb{R}} |f(x)|\, \mu(\mathbb{R}) = \sup_{x\in\mathbb{R}} |f(x)| < \infty$$

by assumption. In particular, let us consider, for a given $t \in \mathbb{R}$, $f_t(x) = 1_{\{x \leq t\}}$: f_t is Borel-measurable and bounded, and

$$\int_{\mathbb{R}} f_t(x)\, d\mu(x) = \int_{-\infty}^{t} d\mu(x) = \mu((-\infty, t]) = F_\mu(t)$$

For discrete and continuous distributions, the Lebesgue integral simply reads:

For μ discrete, $\displaystyle\int_{\mathbb{R}} f(x)\, d\mu(x) = \sum_{x\in C} f(x)\, \mu(\{x\})$. **For μ continuous,** $\displaystyle\int_{\mathbb{R}} f(x)\, d\mu(x) = \int_{\mathbb{R}} f(x)\, p_\mu(x)\, dx$.

1.3 Objects associated to a distribution

The cdf is an example of object associated to a distribution (that moreover characterizes completely the distribution). Here are other examples.

Moments.

Definition 1.3. Let μ be a distribution on $\mathbb{R}$ and $k \geq 0$. If $\int_{\mathbb{R}} |x|^k \, d\mu(x) < \infty$, we then define the *moment of order k* associated to the distribution μ as

$$m_k = \int_{\mathbb{R}} x^k \, d\mu(x)$$

Here are some easy facts:

- As $f(x) = x^k$ is not a bounded function, the moment of order k of a distribution is not always well defined (with the exception of discrete distributions supported on a finite set: all their all moments are always finite).

- If μ has a finite moment of order k, then all its moments of lower order $l \leq k$ are also finite. In general, there is a limiting value k_0 below which all moments are finite and above which all moments are infinite (but k_0 may of course take the value ∞).

- If there exists $C > 0$ such that $\mu([-C, C]) = 1$ (we then say that μ is supported on a compact set), then all the moments of μ are finite and

$$|m_k| \leq \int_{\mathbb{R}} |x|^k \, d\mu(x) = \int_{-C}^{C} |x|^k \, d\mu(x) \leq C^k \int_{-C}^{C} d\mu(x) = C^k$$

There are of course other examples of distributions which are not supported on a compact set and whose moments are all finite (such as e.g. the Gaussian or the log-normal distributions).

In general, an important question is to decide whether a distribution is completely characterized by its moments (which can only possibly happen when all the moments of the distribution are finite). The answer is given by Carleman's theorem.

Theorem 1.4. (Carleman) Let μ be a distribution and $(m_k,\, k \geq 0)$ be the sequence of its moments. If in addition

$$\sum_{k \geq 1} (m_{2k})^{-\frac{1}{2k}} = \infty \tag{3}$$

then the distribution μ is the unique distribution with the sequence of moments $(m_k,\, k \geq 0)$.

Condition (3) is actually a condition on the growth of the moments m_k. It is satisfied in particular if $|m_k| \leq C^k$ for some $C > 0$ (which occurs e.g. for distributions supported on a compact set, as just seen above). Indeed, in this case,

$$m_{2k} \leq C^{2k} \quad \text{so} \quad (m_{2k})^{-\frac{1}{2k}} \geq \frac{1}{C}, \quad \text{so} \quad \sum_{k \geq 1} (m_{2k})^{-\frac{1}{2k}} = \infty$$

More generally, if $|m_k| \leq C \, \exp(k \log k)$, then condition (3) holds. This is the case for example for the Gaussian distribution (in which case $m_k \sim k!$), but not for the log-normal distribution (in which case $m_k \sim \exp(k^2)$)

Stieltjes (or Cauchy) transform.

Definition 1.5. Let μ be a distribution on $\mathbb{R}$ and $z \in \mathbb{C} \backslash \mathbb{R}$. The *Stieltjes transform of* μ is the mapping $g_\mu : \mathbb{C} \backslash \mathbb{R} \to \mathbb{C}$ defined as

$$g_\mu(z) = \int_{\mathbb{R}} \frac{1}{x - z} \, d\mu(z), \quad z \in \mathbb{C} \backslash \mathbb{R}$$

Notice that for $z \in \mathbb{C} \backslash \mathbb{R}$, the function $x \mapsto f_z(x) = \frac{1}{x-z}$ is bounded and continuous on the real line, so $g_\mu(z)$ is always well defined.

Basic properties.

- g_μ is analytic on $\mathbb{C} \backslash \mathbb{R}$
- $\text{Im } g_\mu(z) > 0$ for all $z \in \mathbb{C}$ such that $\text{Im } z > 0$
- $\lim_{v \to \infty} v \, |g_\mu(iv)| = 1$

Moreover, it turns out that any function g satisfying the above three properties is the Stieltjes transform of a distribution μ on $\mathbb{R}$. In addition, we have the following *inversion formula*:

If $a < b$ are continuity points of F_μ, then

$$F_\mu(b) - F_\mu(a) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_a^b \text{Im } g_\mu(u + i\varepsilon) \, du$$

In the case where μ is a continuous distribution with pdf p_μ, the above formula simplifies to

$$p_\mu(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \text{Im } g_\mu(x + i\varepsilon) \quad \forall x \in \mathbb{R}$$

2 Weak convergence of sequences of distributions

Definition 2.1. Let $(\mu_n,\, n \geq 1)$ be a sequence of distributions and μ be another distribution. The sequence μ_n is said to converge weakly to μ as n goes to infinity if

$$\lim_{n\to\infty} F_{\mu_n}(t) = F_\mu(t) \quad \forall t \in \mathbb{R} \text{ continuity point of } F_\mu$$

Notation. $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$.

So weak convergence of distributions means pointwise convergence of the corresponding cdfs, except in the points where the limiting cdf makes a jump[3]. This definition has several equivalents, among which the following, which is part of the so-called portmanteau theorem.

Proposition 2.2. $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$ if and only if for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$,

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\, d\mu_n(x) = \int_{\mathbb{R}} f(x)\, d\mu(x)$$

Checking either of these criteria in order to prove weak convergence is difficult in general. In the sequel, we propose other criteria that are easier to apply, in particular in random matrix theory.

2.1 Various characterizations of weak convergence

Via moments. The full version of Carleman's theorem is given below.

Theorem 2.3. (Carleman) Let $(\mu_n,\, n \geq 1)$ be a sequence of distributions and $(m_k,\, k \geq 0)$ be a sequence of numbers such that for all $k \geq 0$,

$$\lim_{n\to\infty} \int_{\mathbb{R}} x^k d\mu_n(x) = m_k$$

and such that the sequence $(m_k,\, k \geq 0)$ satisfies condition (3). Then $(m_k,\, k \geq 0)$ is a sequence of moments to which corresponds a unique distribution μ, and μ_n converges weakly to μ.

Via Stieltjes transform.

Theorem 2.4. Let $(\mu_n,\, n \geq 1)$ be a sequence of distributions and μ be another distribution. Then $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$ if and only if

$$\lim_{n\to\infty} g_{\mu_n}(z) = g_\mu(z)$$

for all $z \in \mathbb{C}_+ = \{z \in \mathbb{C} : \text{Im } z > 0\}$.

The above two criteria are very useful for (random) matrices, for the following reason. Let $A^{(n)}$ be an $n \times n$ Hermitian matrix, let $\lambda_1^{(n)}, \ldots, \lambda_n^{(n)}$ be its real eigenvalues and let μ_n be the distribution of one of these eigenvalues picked at random. Then

$$m_k^{(n)} = \int_{\mathbb{R}} x^k\, d\mu_n(x) = \frac{1}{n} \sum_{j=1}^n \left(\lambda_j^{(n)}\right)^k = \frac{1}{n} \operatorname{Tr}\left(\left(A^{(n)}\right)^k\right)$$

and

$$g_{\mu_n}(z) = \int_{\mathbb{R}} \frac{1}{x-z}\, d\mu_n(x) = \frac{1}{n} \sum_{j=1}^n \frac{1}{\lambda_j^{(n)} - z} = \frac{1}{n} \operatorname{Tr}\left(\left(A^{(n)} - zI\right)^{-1}\right)$$

In order to compute these quantities, one does actually not need to know what the eigenvalues $\lambda_j^{(n)}$ are!

[3] Notice indeed that asking for pointwise convergence of a sequence of functions towards a limiting discontinuous function in every point of $\mathbb{R}$ would be asking for too much.