Random matrix theory: lecture 12

Weak convergence and moments

Recall: if for $k \geq 0$, $\int_{\mathbb{R}} |x|^k \, d\mu(x) < \infty$ for a given distribution μ,

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

Clearly, not all moments of a distribution are necessarily finite.

Example: let μ be the Cauchy distribution, with pdf

× $$p_\mu(x) = \frac{1}{\pi(1+x^2)} \quad x \in \mathbb{R}; \text{ then } \int_{\mathbb{R}} |x|^k \, d\mu(x) = \infty \quad \forall k \geq 1$$

But even in the case where all moments are finite, the sequence $(m_k)_{k=0}^{\infty}$ does not necessarily characterize entirely the distribution μ.

Carleman's condition:

If $\sum_{k=0}^{\infty} (m_{2k})^{-\frac{1}{2k}} = \infty$, then μ is the only distribution with the sequence of moments $(m_k)_{k=0}^{\infty}$.

Remark: Carleman's condition is a condition on the growth of the moments m_k; it basically requires that $m_k \lesssim e^{k \log k}$.

Example: if $\mu([-c, c]) = 1$ for some $c > 0$, then

$|m_k| \leq \int_{\mathbb{R}} |x|^k \, d\mu(x) = \int_{-c}^{c} |x|^k \, d\mu(x) \leq c^k$ satisfies the condition.

Counter-example: the log-normal distribution μ with pdf

$p_\mu(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left(-\frac{(\log x)^2}{2}\right)$, $x > 0$, has moments $m_k = e^{k^2/2}$.

Proposition

Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of distributions. If there exists a sequence of real numbers $(m_k)_{k=0}^{\infty}$ satisfying Carleman's condition and such that

$$\lim_{n\to\infty}\int_{\mathbb{R}} x^k \, d\mu_n(x) = m_k \quad \forall k \geq 0$$

then there exists a unique distribution μ with the sequence of moments $(m_k)_{k=0}^{\infty}$ such that $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$.

Remarks

- this proposition provides an easy criterion for checking weak convergence of sequences of distributions.
- drawback: in some cases, weak convergence holds even though not all moments are finite.

Weak convergence and Stieltjes transform

Recall: for $z \in \mathbb{C} \setminus \mathbb{R}$, $g_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z} \, d\mu(x)$

× Properties:

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

25.4.07

3

Proposition (inversion formula)

The knowledge of the function $g_\mu(z)$, $z \in \mathbb{C} \setminus \mathbb{R}$ characterizes μ entirely. Moreover, $\forall a < b$,

$$\mu(]a,b[) + \frac{1}{2}\mu(\{a,b\}) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_a^b \operatorname{Im}(g_\mu(u+i\varepsilon))\,du$$

If μ is a continuous distribution with pdf p_μ, then

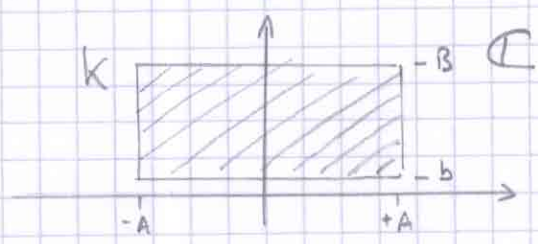
$$p_\mu(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \operatorname{Im}(g_\mu(x+i\varepsilon))$$

Proposition

× $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$ iff $\lim_{n\to\infty} \sup_{z\in K} |g_{\mu_n}(z) - g_\mu(z)| = 0$

$\forall K = [-A, A] \times [b, B] \subset \mathbb{C}$ $\begin{pmatrix} A > 0 \\ B > b > 0 \end{pmatrix}$

"uniform convergence on compacts"



Remark: in practice, we will be satisfied with checking only the condition

$$\lim_{n\to\infty} g_{\mu_n}(z) = g_\mu(z) \quad \forall z \in \mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$$

Relation between all this and matrices?

Let $(A^{(n)})_{n=1}^{\infty}$ be a sequence of Hermitian deterministic matrices of increasing size ($A^{(n)}$ is of size $n \times n$) and let $\lambda_1^{(n)} \dots \lambda_n^{(n)}$ be the (real) eigenvalues of $A^{(n)}$.

We define the following sequence of (discrete) cdf's:

$$F_n(t) := \frac{1}{n} \# \{ j : \lambda_j^{(n)} \le t \} \quad t \in \mathbb{R}$$

and would like to study the (weak) convergence of the sequence $(F_n)_{n=1}^{\infty}$ as $n \to \infty$.

Corresponding distributions μ_n:

In Dirac's notation, $\mu_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j^{(n)}}$

(ie μ_n = uniform distribution on the $\lambda_j^{(n)}$: $\mu_n(\{\lambda_j^{(n)}\}) = \frac{1}{n}$)

Note that $\mu_n(]a, b]) = \frac{1}{n} \# \{ j : a < \lambda_j^{(n)} \le b \} = F_n(b) - F_n(a)$

Now, what is $\int_{\mathbb{R}} f(x) \, d\mu_n(x)$?

We already know that $A^{(n)} = U^{(n)} \Lambda^{(n)} (U^{(n)})^*$ where $U^{(n)}$ is unitary and $\Lambda^{(n)} = \operatorname{diag}(\lambda_1^{(n)}, \dots, \lambda_n^{(n)})$.

For $f : \mathbb{R} \to \mathbb{R}$, let us define $f(A^{(n)}) := U^{(n)} f(\Lambda^{(n)}) (U^{(n)})^*$, with $f(\Lambda^{(n)}) := \operatorname{diag}(f(\lambda_1^{(n)}), \dots, f(\lambda_n^{(n)}))$.

We have: $\frac{1}{n} \operatorname{Tr}(f(A^{(n)})) = \frac{1}{n} \operatorname{Tr}(f(\Lambda^{(n)})) = \frac{1}{n} \sum_{j=1}^{n} f(\lambda_j^{(n)})$

$$= \int_{\mathbb{R}} f(x) \, d\mu_n(x) \; !$$

Moments:

$$m_k^{(n)} = \int_{\mathbb{R}} x^k \, d\mu_n(x) = \frac{1}{n} \sum_{j=1}^{n} \left(\lambda_j^{(n)}\right)^k = \frac{1}{n} \operatorname{Tr}\left(\left(A^{(n)}\right)^k\right) \quad \forall k \geq 0$$

NB: for a given n, the $m_k^{(n)}$ are finite $\forall k \geq 0$

(but they might not converge in the limit $n \to \infty$)

Stieltjes transform:

$$g_n(z) = \int_{\mathbb{R}} \frac{1}{x - z} \, d\mu_n(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\lambda_j^{(n)} - z}$$

$$= \frac{1}{n} \operatorname{Tr}\left(\left(\underbrace{A^{(n)} - z I_n}_{\text{invertible since } z \notin \mathbb{R}} \right)^{-1} \right) \qquad z \in \mathbb{C} \setminus \mathbb{R}$$

Fourier transform:

$$\phi_n(t) = \int_{\mathbb{R}} e^{itx} \, d\mu_n(x) = \frac{1}{n} \sum_{j=1}^{n} e^{i \lambda_j^{(n)} t}$$

$$= \frac{1}{n} \operatorname{Tr}\left(e^{itA^{(n)}} \right) \qquad t \in \mathbb{R}$$

problem: In order to compute the exponential of a matrix, we need to compute either its eigenvalues or all the moments (using the formula: $e^{itA^{(n)}} = \sum_{m=0}^{\infty} \frac{(it)^m}{m!} \cdot \left(A^{(n)}\right)^m$). The Fourier transform is therefore not that useful.

Aside:

Note that $\frac{1}{n} \log \det A^{(n)} = \frac{1}{n} \sum_{j=1}^{n} \log\left(\lambda_j^{(n)}\right) = \frac{1}{n} \operatorname{Tr}\left(\log A^{(n)}\right)$

(assume $A^{(n)} > 0$) $\qquad = \int_{\mathbb{R}_+} \log(x) \, d\mu_n(x)$

Important remark

For a sequence of <u>random</u> matrices $A^{(n)}$, the eigenvalues $\lambda_1^{(n)} \dots \lambda_n^{(n)}$ are random variables; the μ_n's are therefore random distributions, which means that $F_n(t)$, $m_k^{(n)}$, $g_n(z)$ and $\phi_n(t)$ are all random variables. We therefore need to specify in what sense the convergence takes place (i.e. writing $\lim_{n\to\infty} F_n(t) = F_\mu(t)$ is not enough).

Let $(X_n)_{n=1}^{\infty}$ be a generic sequence of random variables.

1) <u>Convergence in expectation</u>: $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$

2) <u>Convergence in probability</u>: (denoted as $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$)

$$\forall \varepsilon > 0, \ \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

3) <u>Almost sure convergence</u>: (denoted as $X_n \xrightarrow[n\to\infty]{} X$ a.s.)

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Note that $3) \Longrightarrow 2) \Longrightarrow 1)$

(provided that the sequence (X_n) is "uniformly integrable")

<u>Special case</u>: convergence to a deterministic limit $X \equiv c$

1) Convergence in expectation simply means $\lim_{n\to\infty} \mathbb{E}(X_n) = c$

2) <u>Proposition</u>: if $\lim_{n\to\infty} \mathbb{E}(X_n) = c$ and $\lim_{n\to\infty} Var(X_n) = 0$,

then $X_n \xrightarrow[n\to\infty]{\mathbb{P}} c$.

<u>Proof</u>: $\mathbb{P}(|X_n - c| > \varepsilon) \leq \dfrac{\mathbb{E}((X_n-c)^2)}{\varepsilon^2}$ (Markov inequality)

$$= \frac{1}{\varepsilon^2} \mathbb{E}\left(\left(X_n - \mathbb{E}(X_n) + \mathbb{E}(X_n) - c\right)^2\right)$$

$$\leq \frac{2}{\varepsilon^2} \Big\{ \underbrace{\mathbb{E}\left((X_n - \mathbb{E}(X_n))^2\right)}_{= Var(X_n) \to 0} + \underbrace{(\mathbb{E}(X_n) - c)^2}_{\to 0} \Big\} \xrightarrow[n\to\infty]{} 0, \quad \forall \varepsilon > 0$$

#

3) <u>Proposition</u>: if $|\mathbb{E}(X_n) - c| = O(\frac{1}{n})$ and $Var(X_n) = O(\frac{1}{n^2})$,

then $X_n \xrightarrow[n\to\infty]{} c$ a.s.

<u>Proof</u>: By the same sequence of inequalities as above,

we obtain that $\mathbb{P}(|X_n - c| > \varepsilon) = O(\frac{1}{n^2})$ $\forall \varepsilon > 0$,

so $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - c| > \varepsilon) < \infty$ $\forall \varepsilon > 0$.

The Borel-Cantelli lemma thus implies that

$\mathbb{P}(|X_n - c| > \varepsilon$ infinitely often$) = 0$ $\forall \varepsilon > 0$

which is an equivalent condition for the

almost sure convergence of the sequence (X_n) towards c.

#

A famous example of (weak) convergence of a sequence of random distributions

- Let $(X_n)_{n=1}^{\infty}$ be a sequence of i.i.d. random variables with distribution μ (i.e. $\mathbb{P}(X_n \in B) = \mu(B)$ $\forall B \in \mathcal{B}(\mathbb{R})$, $\forall n \geq 1$)
- Let $F_n(t) := \frac{1}{n} \cdot \# \{ j : X_j \leq t \}$, $t \in \mathbb{R}$. That is, F_n is the cdf of the empirical distribution $\mu_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j}$ of the first n random variables.
- Note that $F_n(t)$ is a random variable for a given n & t, so that μ_n is a random distribution for a given n.

Proposition: (convergence of the empirical distribution)

$$\lim_{n \to \infty} F_n(t) = F_\mu(t) \quad \text{a.s.}, \ \forall t \in \mathbb{R}$$

Proof: For a given $t \in \mathbb{R}$, let $Y_j := 1_{\{X_j \leq t\}}$ i.i.d. random var.

$F_n(t) = \frac{1}{n} \sum_{j=1}^{n} Y_j$, so by the law of large numbers,

$$\lim_{n \to \infty} F_n(t) = \mathbb{E}(Y_1) \quad \text{a.s.}$$

Since $\mathbb{E}(Y_1) = \mathbb{P}(X_1 \leq t) = F_\mu(t)$,

the proposition is proved. #