CPSC 664, Spring 2023 | Scribe: Peixin You
Feb. 7, 2023 | Lecturer: Tim Kunisky

Lecture 7: Moment Problems

1 Problem Statement

This is a very classical topic in probability theory. In general, this asks when knowing the moments of a probability distribution or limits of moments of a sequence of distributions will help us identify the underlying distribution or its limit.

Formally,

1. The simplest version of this kind of question is: for a given $\{m_k\}_{k\geq 1}$ we want to know the **existence**: if there exists some probability measure μ over $\mathbb{R}$ such that

$$\mathbb{E}_{x\sim\mu} X^k = m_k, \forall k$$

   We quickly recall that, in a more analysis notation, we could also write $\mathbb{E}_{x\sim\mu} X^k$ as $\int x^k d\mu(x)$. When μ has a density that we denote as p, we could rewrite the integral to $\int x^k p(x)dx$.

   After asking if there exists such a μ, another natural question we would like to answer is **uniqueness:** when such μ exists, will it be unique or not?

2. We have seen a more advanced question in the number partitioning lecture: given random variables X_n and for every fixed k the moment

$$\lim_{n\to\infty} \mathbb{E}X_n^k \to \int x^k d\mu(x)$$

   then the question is when we could conclude $X_n \to \mu$ in some other sense of convergence.

3. At the end of this lecture, we want to discuss the appearances in random matrix theory of convergence in moments.

2 Existence and uniqueness

So let's start with the simplest version:

Given the sequence of moments of some μ, $m_k = \int x^k d\mu(x)$. We want to know how and when we could recover μ from the given information. Part of this question is like an algorithmic question when we give you these numbers, how do we calculate μ? More abstractly, we want to clarify if there exists a $\mu' \neq \mu$ such that $m_k = \int x^k d\mu'(x)$.

To make our life easier, we assume our μ has density function $p(x)$; this allows us to compute μ by integrating p.

Now we first introduce *characteristic function*.

Definition 2.1. Characteristic function $\Phi(t) \in \mathbb{C}$ *is defined as*

$$\Phi(t) = \mathop{\mathbb{E}}_{X\sim\mu} e^{itX} = \int_{-\infty}^{\infty} e^{itx} p(x) dx.$$

Example 2.2. *The characteristic function of the Gaussian law* $\mathcal{N}(\mu, \sigma^2)$ *is* $\mathsf{exp}\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$. *The characteristic function of the exponential law* $\mathsf{Exp}(\lambda)$*, which we saw appear in the number partitioning lecture, is* $\frac{1}{1-\frac{i}{\lambda}t}$.

Remark 2.3. *One way to think of this is just like a moment-generating function but evaluated with complex parameters.*

Theorem 2.4. *Given a* Φ *with some nice properties, for instance* Φ *is in* $\mathcal{L}^1$*, we have*

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \Phi(t) dt.$$

Proof. Notice that

$$RHS = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{it(y-x)} p(y) dy dt.$$

The idea of calculating this integral is to integrate $e^{it(y-x)}$ as a delta function.

To prove this more formally, we can show the result if p is some Gaussian, and then approximate other p by some linear combination of Gaussian densities. Then we will use that the map $p \mapsto \Phi$, and the inversion map are both linear. So we can pass the linear combination through the integral and make some arguments. □

Now we see if we know $\Phi(t)$ on $t \in \mathbb{R}$, usually it's enough to recover p.

We would like to move to the next question now: how do the moments determine Φ? The main idea of solving this question is using Taylor expansion.

Now let's calculate the derivatives of Φ:

$$\Phi^{(k)}(t) := \frac{d^k \Phi}{dt^k} = \frac{d^k}{dt^k} \int_{-\infty}^{\infty} e^{itx} p(x) dx = \int_{-\infty}^{\infty} e^{itx} (ix)^k p(x) dx.$$

At $t = 0$, we have $\Phi^{(k)}(0) = i^k m_k$ so the Taylor series will be

$$\sum_{k=0}^{\infty} \frac{i^k m_k}{k!} t^k.$$

In the best case, we hope the Taylor series has infinite radius of convergence. So the series will give us the characteristic function, which in turn gives us μ. If $\frac{|m_k|^{1/k}}{k} \to 0$ then

$$\left|\frac{m_k}{k!}\right| \leq \frac{|m_k|}{(k/e)^k} \leq \left(\frac{|m_k|^{1/k} e}{k}\right)^k \leq \varepsilon^k$$

for every sufficiently large k and for all $\varepsilon > 0$. Then we know the Taylor series

$$\Phi(t) = \sum_{k=0}^{\infty} \frac{i^k m_k}{k!} t^k$$

converges for all t, so we could recover $p(x)$.

It is also possible to weaken this condition. If $\frac{|m_k|^{1/k}}{k}$ is bounded, or equivalently, $|m_k| \leq CM^k k!$ where C, M are constant, then we have

$$\left|\frac{m_k}{k!}\right| \leq \frac{|m_k|}{(k/e)^k} \leq \left(\frac{|m_k|^{1/k} e}{k}\right)^k \leq \Theta\left(\frac{1}{M}\right)^k$$

for every sufficient large k. This gives convergence on $\left(-\Theta\left(\frac{1}{M}\right), \Theta\left(\frac{1}{M}\right)\right)$.

Now let's look at the Taylor series at other center t_0:

$$\sum_{k=0}^{\infty} \frac{\Phi^{(k)}(t_0)}{k!}(t-t_0)^k.$$

Notice that

$$\left|\Phi^{(k)}(t)\right| \leq \int_{-\infty}^{\infty} |x|^k p(x) dx$$

which is like the moment m_k but with an absolute value. We denote it as $\tilde{m}_k$. We can bound these by the even moments $m_{2\ell}$.

Remark 2.5. *We have:*

$$\tilde{m}_{2k} = m_{2k}$$
$$\tilde{m}_{2k+1} \leq m_{2k+2}^{\frac{2k+1}{2k+2}}$$

With the inequality above, we have

$$\tilde{m}_{2k}^{1/(2k)} \leq m_{2k}^{1/(2k)}$$
$$\tilde{m}_{2k+1}^{1/(2k+1)} \leq m_{2k+2}^{1/(2k+2)}$$

This means if the m_k is bound, that is actually what we assumed, then the absolute value will behave the same.

We then find that the Taylor series converges on a ball of radius $\Theta(1/M)$ around any $t_0 \in \mathbb{R}$. A complex analysis argument using analytic continuation then shows that the moments determine the characteristic function on $\mathbb{R}$.

Theorem 2.6. *If $|m_k| \leq CM^k k!$ for some constant C, M, then μ is determined uniquely by the moments m_k.*

Let's see some examples on which this theorem works.

Example 2.7. *Theorem 2.6 applies to any compactly supported distribution, and Gaussian distribution, any sub-Gaussian distribution, any exponential distribution, and any Poisson distribution.*

Also, let's see one example that doesn't satisfy the condition of Theorem 2.6.

Example 2.8 (Log-normal distribution). *$\mu = e^{\mathcal{N}(0,1)}$ having density $p(x)$. In this case, $m_k = e^{k^2/2}$ which is not good, since what we allowed is $k!$ which is $k^{O(k)}$ or $e^{O(k \log k)}$. Actually, we can check that we have the same moments for our distribution with density $p(x)$ and another distribution with density proportional to $p(x)(1 + \delta \sin(2\pi \log x))$ for small $\delta > 0$.*

2.1 Moment sequences

Let's look at one interesting side question. Before, we had m_k, and we were promised that they are the moment of some distribution and tried to find that distribution. Now we want to go one step further; suppose $m_k \in \mathbb{R}$ are arbitrary numbers. We want to ask if there exists any distribution μ such that $m_k = \int x^k d\mu(x)$.

The answer is not always; a simple condition we need to satisfy is m_{2k} needs to be larger or equal to 0. More generally, for all polynomial $p \in \mathbb{R}[x]$, we need to have $\int p(x)^2 d\mu(x) \geq 0$, which may be expanded to a linear inequality condition on the m_k. In fact, these are the only necessary conditions.

Theorem 2.9. *Suppose for all k we have*

$$M := \begin{bmatrix} m_0 = 1 & m_1 & m_2 & \cdots & m_k \\ m_1 & m_2 & m_3 & \cdots & \cdots \\ m_2 & m_3 & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & m_{2k-1} \\ m_k & \cdots & \cdots & m_{2k-1} & m_{2k} \end{bmatrix} \succeq 0,$$

i.e., $v^T M v \geq 0$ for all v. One can check this condition is equivalent to

$$\int (v_0 + v_1 x + \cdots v_k x_k)^2 d\mu(x) \geq 0.$$

Then there exists a μ such that $m_k = \int x^k d\mu(x)$.

3 Convergence of moments

Let's first recall our second question: suppose we have sequences of moments converge to some number

$$\lim_{n\to\infty} \mathbb{E}X_n^k \to m_k = \int x^k d\mu(x)$$

for all k. We want to know if this implies random variable X_n converge in distribution to some law μ, i.e.,

$$\mathbb{P}[X_n \leq t] \to \mu((-\infty, t])$$

We will answer this question in two steps. Let μ_n be the law of X_n. First, let's consider the simplest case in which the implication holds:

Theorem 3.1. *Suppose there exists an $M > 0$ such that*

- $X_n \in [-M, M]$ *with probability* 1
- $\mathsf{supp}(\mu) \subseteq [-M, M]$

Then we have

$$\mathbb{P}[X_n \leq t] = \int_{-M}^{M} \mathbb{1}_{[-M,t]}(x) d\mu_n(x) \to \int_{-M}^{M} \mathbb{1}_{[-M,t]}(x) d\mu(x)$$

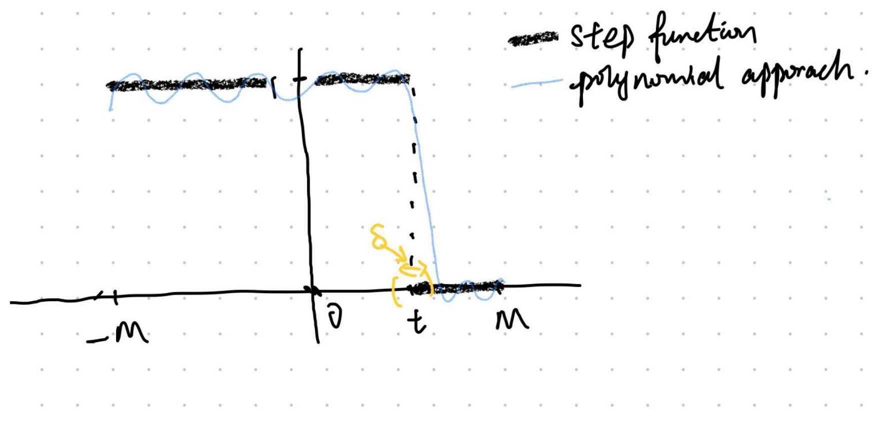
Proof. Let's denote μ_n be the law of X_n, then the assumption is equivalent to

$$\int x^k d\mu_n(x) \to \int x^k d\mu(x)$$

Then we want to show

$$\mathbb{P}[X_n \leq t] = \int_{-M}^{M} \mathbb{1}_{[-M,t]}(x) d\mu_n(x) \to \int_{-M}^{M} \mathbb{1}_{[-M,t]}(x) d\mu(x)$$

We want to find a polynomial p that approximates $\mathbb{1}_{[-M,t]}(x)$ to additive error ε away from t, additive error for instance let's say 2 near t.



Then we could check that

$$\left|\int_{-M}^{M} \mathbb{1}_{[-M,t]}(x) d\mu_n(x) - \int_{-M}^{M} \mathbb{1}_{[-M,t]}(x) d\mu(x)\right| \leq \left|\int p(x) d\mu(x) - \int p(x) d\mu(x)\right| + 4\varepsilon M$$
$$+ 2(\mu_n([t-\delta, t+\delta]) + \mu([t-\delta, t+\delta]))$$

We could choose ε, δ small, and choose p, allowing us to make all terms arbitrarily small. □

More generally, we can use the following theorem.

Theorem 3.2 (Levy). *A random variable X_n converge to distribution μ if and only if the characteristic function $\Phi_{X_n}(t)$ converge to the characteristic function $\Phi_\mu(t)$ for all $t \in \mathbb{R}$.*

Corollary 3.3. *Under the same growth condition as before, i.e., $|m_k| \leq CM^k k!$, if we have*

$$\mathbb{E}X_n^k \to \int x^k d\mu(x) = m_k,$$

then X_n converge in distribution to μ.

Example 3.4 (Central limit theorem). *Let $\mu = \mathcal{N}(0,1)$, then we could see $m_k = 0$ when k is odd, $m_k = (k-1)!!$ when k is even. With this, we can see $m_k \le (k-1)!! \le k!$ satisfies the growth condition.*

Let's define

$$A_1, \dots, A_n \overset{iid}{\sim} \mathsf{Unif}(\{\pm 1\}), \quad X_n = \frac{A_1 + \cdots + A_n}{\sqrt{n}}$$

Let's assume k is even and look at

$$\begin{aligned} \mathbb{E}X_n^k &= n^{-k/2} \sum_{i_1,\dots,i_k} \mathbb{E}[A_{i_1} \cdots A_{i_k}] \\ &= n^{-k/2} \#\{(i_1,\dots,i_k) \in [n]^k : \textit{each index happens even number of times}\} \\ &\approx n^{-k/2}(n) \cdot (n-1) \cdot \cdots \cdot (n-k/2+1) \#\{\textit{pairing of } 1,\dots,k\} \end{aligned}$$

Since n goes to infinity and k is constant, we could say $n^{-k/2}(n) \cdot (n-1) \cdot \cdots \cdot (n-k/2+1)$ cancel each other approximately. So we have

$$\mathbb{E}X_n^k \to \#\{\textit{pairing of } 1,\dots,k\} = \mathbb{1}\{k \textit{ even}\}(k-1)!!.$$

With some more careful argument, we could prove the central limit theorem for more general distributions of A_i in this way.

Example 3.5. *Similar, though more complicated, versions of the same ideas are used to derive the $\mathsf{Exp}(\sqrt{(2\pi)/3})$ limit theorem for spacings in the random number partitionining problem.*

4 Appearance in random matrix theory

Now we want to talk about the relevance to random matrices. Let's fix some notation here; in the following, $A_n \in \mathbb{R}^{n \times n}_{\text{sym}}$ are symmetric random matrices. Since they are symmetric, they have real eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{R}$.

We define μ_n the so-called *empirical spectral distribution* of A_n as follows:

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}.$$

An important thing we need to keep in mind is these are *random* probability measures.

We want to ask in what sense can we have μ_n converges to a μ, which is a deterministic probability measure?

These random probability measures are pretty complicated, but we have many tools to discuss the convergence of scalar random variables. So, let's build random variables from μ_n,

$$\mu_n([a,b]) = \frac{\#\{i \in [n] : \lambda_i \in [a,b]\}}{n}.$$

We can then ask if this converges in some standard probability sense to $\mu([a,b])$, for all choices of a, b.

Notice that $\int x^k d\mu_n(x)$ is a random variable instead of a number. We will have convenient tools to calculate the expectation of this random variable:

$$\mathop{\mathbb{E}}_{A_n} \int x^k d\mu_n(x) = \mathop{\mathbb{E}}_{A_n} \frac{1}{n} \sum_{i=1}^{n} \lambda_i^k = \frac{1}{n} \mathop{\mathbb{E}}_{A_n} \mathsf{Tr}(A_n^k).$$

Now we want to ask if this converges to $\int x^k d\mu(x)$. This is a necessary condition to have the above kind of convergence, but generally is not enough.

This technique will only tell us the expectation $\mathbb{E}\mu_n([a,b])$ converge to $\mu([a,b])$, but what we want is convergence in probability, which asks if can we get the probability that fraction differs from the right-hand side by any ε goes to zero. We can't get this by only using the expectations of moments. One additional piece of information that suffices is enough control over $\mathsf{Var}(\frac{1}{n}\mathsf{Tr}(A_n^k))$. With additional such conditions, we could prove the convergence in probability by using Chebyshev inequality.