STATS310A - Lecture 19

Persi Diaconis
Scribed by Michael Howes

12/02/21

Contents

1 Fourier inversion

Theorem 1. *Suppose that μ is a probability on $\mathbb{R}$ with charateristic function ϕ. If $\mu(\{a\}) = \mu(\{b\}) = 0$, then*

$$\mu((a,b)) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi(t)dt. \tag{1}$$

To prove this we will need several classic calculus facts. We won't prove them here but their derivations are in Section 26 of Billingsley.

(a) Let $S(T) = \int_0^T \frac{\sin(t)}{t}dt$ ($S(T)$ is also known as $\mathrm{sinc}(T)$). While $\int_0^\infty \frac{\sin(t)}{t}dt$ does not exist as a Lebesgue integral, a classical calculation shows that

$$\lim_{T \to \infty} S(T) = \frac{\pi}{2}.$$

Thus $S(T)$ is bounded on $[0, \infty)$. Let sgn be the function

$$\mathrm{sgn}(\theta) = \begin{cases} -1 & \text{if } \theta < 0, \\ 0 & \text{if } \theta = 0, \\ 1 & \text{if } \theta > 0. \end{cases}$$

By using a change of variables, one can show that

$$\int_0^T \frac{\sin(t\theta)}{t}dt = \mathrm{sgn}(\theta)S(T|\theta|). \tag{2}$$

(b) Using Taylor's theorem we have for all real t.

i. $\left|e^{ix} - 1\right| \le \min\{|x|, 2\}.$

ii. $\left|e^{ix} - (1 + ix)\right| \le \min\left\{\frac{1}{2}x^2, 2|x|\right\}.$

Using (ii.) we see that the integrand in equation (1) converges to $b - a$ as $t \to 0$. Thus the integrand in (1) is continuous and by (i.) it is bounded. Thus the integral in (1) exists for every T.

With these calculus facts we are ready to prove Theorem 1.

Proof. For $T > 0$, let

$$I_T = \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt.$$

Note that

$$I_T = \frac{1}{2\pi} \int_{-T}^{T} \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \mu(dx) dt.$$

The above integrand is a bounded function and the space $[-T, T] \times \mathbb{R}$ has finite measure with respect to $dt \times \mu(dx)$. Thus we can apply Fubinni's theorem which gives

$$I_T = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt \mu(dx).$$

We will now use equation (2) and the fact sin and cos are odd and even respectively. This gives us

$$I_T = \frac{1}{2\pi} \int_{\mathbb{R}} 2 \int_{0}^{T} \frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t} dt$$

$$= \int_{\mathbb{R}} \left[\frac{\operatorname{sgn}(x-a)}{\pi} S(T|x-a|) - \frac{\operatorname{sgn}(x-b)}{\pi} S(T|x-b|) \right] \mu(dx).$$

The function $f_T(x) = \frac{\operatorname{sgn}(x-a)}{\pi} S(T|x-a|) - \frac{\operatorname{sgn}(x-b)}{\pi} S(T|x-b|)$ is uniformly bounded over x and T. We know that for $x \ne a, b,$

$$\frac{1}{\pi} S(T|x-a|), \frac{1}{\pi} S(T|x-b|) \to \frac{1}{\pi} \cdot \frac{\pi}{2} = \frac{1}{2}.$$

Thus if $\operatorname{sgn}(x-a) = \operatorname{sgn}(x-b)$ and $x \ne a, b$, then $f_T(x) \to 0$. Also if $\operatorname{sgn}(x-a) \ne \operatorname{sgn}(x-b)$ and $x \ne a, b$, then $f_T(x) \to 1$. If $x = a$, then $\operatorname{sgn}(x-b) = -1$ and $f_T(x) \to \frac{1}{2}$. Lastly, if $x = b$, then $\operatorname{sgn}(x-a) = 1$ and $f_T(x) \to \frac{1}{2}$. Summarising this, we can conclude that f_T converges pointwise to the function $\psi_{a,b}$ where

$$\psi_{a,b}(x) = \begin{cases} 0 & \text{if } x < a, \\ \frac{1}{2} & \text{if } x = a, \\ 1 & \text{if } a < x < b, \\ \frac{1}{2} & \text{if } x = b, \\ 0 & \text{if } x > b. \end{cases}$$

Since $\mu(\{a, b\}) = 0$, we can conclude that f_T converges to $\delta_{(a,b)}$ μ-almost surely. Thus, by the dominated convergence theorem we have

$$\lim_{T \to \infty} I_T = \lim_{T \to \infty} \int_{\mathbb{R}} f_T(x) \mu(dx) = \int_{\mathbb{R}} \delta_{(a,b)}(x) \mu(dx) = \mu((a, b)). \qquad \square$$

Note that this proof also shows that

$$\mu((a, b)) + \frac{1}{2} \mu(\{a, b\}) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt,$$

which was the version of the Fourier inversion theorem stated on Tuesday.

Remark 1. To be honest, while careful, this proof "sort of stinks". It doesn't give any feeling for what's going on. We'll come back to this later when we look at a discrete version of characteristic functions. This "lack of transparency" is what led Stein to develop Stein's method. It's also why we studied Lindeberg's proof of the central limit theorem in class.

Remark 2. Note that if F is the cumulative distribution function of μ, then equation (1) can be written as

$$F(b) - F(a) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt.$$

In particular for $c \in \mathbb{R}$ and $h > 0$,

$$\frac{F(c+h) - F(c)}{h} = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itc}(1 - e^{-ith})}{ith} \phi(t) dt.$$

If we assume that ϕ is integrable (so that $\int_{\mathbb{R}} |\phi(t)| dt < \infty$), then an analysis of the above equation can be used to show that μ has a density f given by

$$f(c) = F'(c) = \int_{\mathbb{R}} e^{-itc} \phi(t) dt.$$

Example 1 (The usual CLT). The following is a typical application. Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$ and characteristic function ϕ. Let $S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j$. Since the first two moments of X_j exist ϕ is twice continuously differentiable with $\phi'(0) = 0$ and $\phi''(0) = -1$.

Since $X_1, X_2, \ldots,$ are independent, we know the that

$$\phi_{S_n}(t) = \phi(t/\sqrt{n})^n = \left(1 - \frac{t^2}{2n} + o\left(1/n\right)\right)^n \to e^{-\frac{t^2}{2}} = \phi_{\mathcal{N}}(t),$$

where $\phi_{\mathcal{N}}$ denotes the characteristic function of the standard normal distribution. A careful justification thus again follows from Taylor's theorem. See Billingsley 26.4 for the details (it really is just Taylor's theorem).

2   Fourier analysis on C_n

Fourier analysis is so nice and so useful that it is good to see it in different contexts. To "make it our friend", let's look at Fourier analysis on an important finite space.

Let $C_n = \{0, 1, \ldots, n-1\}$ denote the group of integers modulo n. We can think of C_n as n points on a circle. A probability on C_n is simply a function $\mu : C_n \to \mathbb{R}$ such that $\mu(j) \geq 0$ and $\sum_{j=0}^{n-1} \mu(j) = 1$. For example we can take $\mu(-1) = \mu(1) = \frac{1}{2}$ and $\mu(j) = 0$ for all other j. The measure μ corresponds to a "drunkard's walk."

The convolution of two probability μ and ν on C_n is defined to be

$$(\mu * \nu)(j) = \sum_{k=0}^{n-1} \mu(k) \nu(j - k).$$

This corresponds to picking k_1 and k_2 from μ and ν independently and then asking what is the probability that $k_1 + k_2 = j$. For a probability μ we can inductively define

$$\mu^{*l} = \mu * \mu^{*(l-1)}.$$

Let $U(j) = \frac{1}{n}$ be the uniform distribution on C_n. We wish to use Fourier analysis to prove the following.

Theorem 2. *Suppose that n is odd and μ is the probability on C_n given by $\mu(-1) = \mu(1) = \frac{1}{2}$. Then $\mu^{*l}(j) \to U(j)$ for all j.*

Note that if n is even, then we have a parity problem. The measure $\mu^{*(2l)}$ is supported on the even numbers in $\{0, 1, \ldots, n-1\}$ and $\mu^{*(2l-1)}$ is supported on the odd numbers in $\{0, 1, \ldots, n-1\}$ thus we will not get convergence. To prove Theorem 2 we will need discrete Fourier analysis.

Definition 1. *For a function $f : C_n \to \mathbb{R}$, define the Fourier transform of f to be $\widehat{f} : C_n \to \mathbb{C}$, given by*

$$\widehat{f}(j) = \sum_{k=0}^{n-1} e^{2\pi i jk/n} f(k).$$

One can check that

(a) For all $f_1, f_2 : C_n \to \mathbb{R}$, $\widehat{f_1 * f_2}(j) = \widehat{f_1}(j) \cdot \widehat{f_2}(j)$.

(b) If $U(j) = \frac{1}{n}$ for all j, then

$$\widehat{U}(j) = \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i jk/n} = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j \neq 0. \end{cases}$$

(c) If $\mu(-1) = \mu(1) = \frac{1}{2}$ and $\mu(j) = 0$ otherwise, then

$$\mu(j) = \frac{1}{2} e^{-2\pi i j/n} + \frac{1}{2} e^{2\pi i j/n} = \cos\left(\frac{2\pi j}{n}\right).$$

We have an inversion theorem for the discrete Fourier transform, just like for characteristic functions on $\mathbb{R}$.

Theorem 3. *For any function $f : C_n \to \mathbb{R}$, we have*

$$f(k) = \frac{1}{n} \sum_{j=0}^{n-1} e^{-2\pi i jk/n} \widehat{f}(j).$$

Proof. Note that both sides of the above equation are linear in f. Since the space of all $f : C_n \to \mathbb{R}$ is a finite dimensional vector space it suffices to prove the above equation for a basis of our choice. The functions

$$\delta_a(k) = \begin{cases} 1 & \text{if } k = a, \\ 0 & \text{if } k \neq a. \end{cases}$$

are the basis we will work with. Note that

$$\widehat{\delta_a}(j) = \sum_{k=0}^{n-1} e^{2\pi i jk/n} \delta_a(k) = e^{2\pi i ja/n}.$$

Thus we have

$$\frac{1}{n} \sum_{j=0}^{n-1} e^{-2\pi i jk/n} \widehat{f}(j) = \frac{1}{n} \sum_{j=0}^{n-1} e^{-2\pi i jk/n} e^{2\pi i ja/n}$$

$$= \frac{1}{n} \sum_{j=0}^{n-1} e^{-2\pi i j(a-k)/n}$$

$$= \begin{cases} 1 & \text{if } k = a, \\ 0 & \text{else.} \end{cases}$$

$$= \delta_a(k). \qquad \square$$

Thus we can prove Theorem 2.

Proof. Since n is odd, $\left|\cos\left(\frac{2\pi j}{n}\right)\right| < 1$ for $j \in C_n\backslash\{0\}$. This is because $\cos\left(\frac{2\pi j}{n}\right) = \pm 1$ and $0 \le j \le n-1$ implies $j = 0$ or $j = \frac{n}{2}$. Thus for every $j \in C_n$ we have

$$\widehat{\mu^{*l}}(j) = \widehat{\mu}(j)^l = \cos\left(\frac{2\pi j}{k}\right)^l \to \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j \neq 0. \end{cases} = \widehat{U}(j).$$

The discrete Fourier inversion formula thus given $\mu^{*l}(j) \to U(j)$ for all k. $\qquad\qquad \square$

If one uses the approximation $\cos\left(\frac{2\pi j}{n}\right) = 1 - \frac{2\pi^2 j^2}{n^2} + O(1/n^4)$, then some careful calculus gives a quantitative version of Theorem 2.

Theorem 4. *For n odd and μ and U as in Theorem 2,*

$$\frac{1}{6}e^{-\frac{\pi^2}{2}l/n^2} \le \left\|\mu^{*l} - U\right\|_{TV} \le 6e^{-\frac{\pi^2}{2}l/n^2}.$$

That is, it takes order n^2 for the walk to get random and n^2 is both necessary and sufficient.

This story extends to any finite or compact group G. If μ is a probability that is not supported on a proper coset of G, then μ^{*l} converges in distribution to the uniform measure on G. See Persi's book "Group representations in probability and statistics" (freely available on his website).

3   Error bounds

Fourier analysis can be used to give error bounds in the central limit theorem. For instance, the following can be proved with Fourier analysis,

Theorem 5 (Berry-Esseen Theorem). *Let $\{X_i\}_{i=1}^\infty$ be i.i.d. with mean 0, variance σ^2 and finite third moment $\gamma = \mathbb{E}[|X_i|^3]$. If $S_n = X_1 + \ldots + X_n$, then*

$$\sup_{x \in \mathbb{R}} \left|\mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) - \Phi(x)\right| \le \frac{C\gamma}{\sqrt{n}\sigma^3}.$$

Where C is a constant that does not depend on X_i and $C \le 0.4748\ldots$.

It is known that there is a choice of X_i that is supported on two points for which $C > 0.40873$ is needed.

There are also multivariate versions (where the supremum is over a class of sets such as the class of all convex sets). Most of the super sharp bounds are proved by Fourier analysis although there are also Stein's method proofs with good constants. These can be used for sums of dependent random variables. The subject is still in active development. Some good references for Stein's method and the central limit theorem are:

- "Normal approximation by Stein's method" by Chen, Goldstein and Shao.

- "High-dimensional Central Limit Theorems by Stein's method" by Fang and Koike.

- "STATS 324: Stein's method" - a course running in the spring by Sourav Chatterjee on Stein's method.

On the other hand, a $\frac{c}{\sqrt{n}}$ error can be pretty disappointing and people often resort to simulation or correction terms to get better convergence.

3.1 Edgeworth corrections

Edgeworth corrections are corrections of the form

$$\mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) + \frac{H_3(x)}{\sqrt{n}} + O(1/n),$$

where H_3 is the 3rd Hermite polynomial. These can make a big difference when n is 'small' (eg $n = 100$). To get Berry-Esseen bounds for this type of approximation is hard honest work. The book "Normal approximation and asymptotic expansions" by Bhattachrya and Ranga Rao has details. Another type of corrections are "Daniel's type aprroximations". See "Small sample asymptotics" by Chris Field. Most of all the above uses Fourier techniques.

4 Course summary

Today is the last lecture and the final exam will be on Tuesday December 7. The final will cover the entire course but will emphasize topics covered after the midterm. The post midterm topics are:
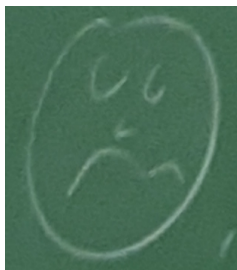
- Covergence in probability weak convergence, almost sure convergence.

- Strong law of large numbers.

- Total variation distance.

- Poisson heuristic, Stein's identity, Stein's method for proving Poisson convergence, dependency graphs, examples (law of small numbers, birthday problem, coupon collector and matching problem).

- Central limit theorem - normal approximation, Lindeberg's condition, Lyapunov's theorem, examples.

- Characteristic functions, basic properties (continuity, behaviour under convolutions, portmanteau theorem, uniqueness, inversion, basic examples).

There will be more or less one question from each of the dot points above. At the moment the exam has some choice. Choose 3 out of 4 "state and prove" type theorems, choose 3 out of 4 "problems" (easier than homework problems). So about half an hour per problem and six problems in total.

You may bring one side $8\frac{1}{2} \times 11$ inch sheet of notes with notes handwritten on one side – to be handed in. The exam is intended to be straight forward.

4.1 Additional topics

There are topics which are important but weren't covered in class.

Three in particular are:

- Kolmogorov's 0-1 law.

- Uniform integrability (when do we have $\mathbb{E}[X_n] \to \mathbb{E}[X]$?).

- Large deviations (convergence in the tails).

The book does a good job covering these and none of them are particularly hard but they should be in our tool kits. Persi would have also liked to have covered infinitely divisible laws. The book has a succinct and good treatment. Despite there being some topics we didn't cover, we did cover a lot and Persi is happy with where we got. Those taking 310B are in for a great course with Sourav Chatterjee.