Lecture 2

Introduction to Some Convergence theorems

Friday 14, 2005
Lecturer: Nati Linial
Notes: Mukund Narasimhan and Chris Ré

2.1 Recap

Recall that for $f : \mathbb{T} \to \mathbb{C}$, we had defined

$$\hat{f}(r) = \frac{1}{2\pi} \int_{\mathbb{T}} f(t) e^{-irt} \, dt$$

and we were trying to *reconstruct* f from $\hat{f}$. The classical theory tries to determine if/when the following is true (for an appropriate definition of equality).

$$f(t) \stackrel{??}{=} \sum_{r \in \mathbb{Z}} \hat{f}(r) e^{irt}$$

In the last lecture, we proved Fejér's theorem $f * k_n \to f$ where the $*$ denotes convolution and k_n (Fejér kernels) are trignometric polynomials that satisfy

1. $k_n \geq 0$
2. $\int_{\mathbb{T}} k_n = 1$
3. $k_n(s) \to 0$ uniformly as $n \to \infty$ outside $[-\delta, \delta]$ for any $\delta > 0$.

If X is a finite abelian group, then the space of all functions $f : X \to \mathbb{C}$ forms an algebra with the operations $(+, *)$ where $+$ is the usual pointwise sum and $*$ is convolution. If instead of a finite abelian group, we take X to be $\mathbb{T}$ then there is no unit in this algebra (i.e., no element h with the property that $h * f = f$ for all f). However the k_n behave as *approximate units* and play an important role in this theory. If we let

$$S_n(f, t) = \sum_{r=-n}^{n} \hat{f}(r) e^{irt}$$

Then $S_n(f, t) = f * D_n$, where D_n is the Dirichlet kernel that is given by

$$D_n(x) = \frac{\sin\left(n + \frac{1}{2}\right) s}{\sin \frac{s}{2}}$$

The Dirichlet kernel does not have all the nice properties of the the Fejér kernel. In particular,

1. D_n changes sign.
2. D_n does not converge uniformly to 0 outside arbitrarily small $[-\delta, \delta]$ intervals.

Remark. The choice of an appropriate kernel can simplify applications and proofs tremendously.

2.2 The Classical Theory

Let G be a locally compact abelian group.

Definition 2.1. A character on G is a *homomorphism* $\chi : G \to \mathbb{T}$. Namely a mapping satisfyin $\chi(g_1+g_2) = \chi(g_1)\chi(g_2)$ for all $g_1, g_2 \in G$.

If χ_1, χ_2 are any two characters of G, then it is easily verified that $\chi_1\chi_2$ is also a character of G, and so the set of characters of G forms a commutative group under multiplication. An important role is played by $\hat{G}$, the group of all continuous characters. For example, $\hat{\mathbb{T}} = \mathbb{Z}$ and $\hat{\mathbb{R}} = \mathbb{R}$.

For any function $f : G \to \mathbb{C}$, associate with it a function $\hat{f} : \hat{G} \to \mathbb{C}$ where $\hat{f}(\chi) = \langle f, \chi \rangle$. For example, if $G = \mathbb{T}$ then $\chi_r(t) = e^{irt}$ for $r \in \mathbb{Z}$. Then we have $\hat{f}(\chi_r) = \hat{f}(r)$. We call $\hat{f} : \hat{G} \to \mathbb{C}$ the Fourier transform of f. Now $\hat{G}$ is also a locally compact abelian group and we can play the same game backwards to construct $\hat{\hat{f}}$. Pontryagin's theorem asserts that $\hat{\hat{G}} = G$ and so we can ask the question: Does $\hat{\hat{f}} = f$? While in theory Fejér answered the question of when $\hat{f}$ uniquely determines f, this question is still left unanswered.

For the general theory, we will also require a normalized nonnegative measure μ on G that is translation invariant: $\mu(S) = \mu(a+S) = \mu(\{a+s \,|\, s \in S\})$ for every $S \subseteq G$ and $a \in G$. There exists a unique such measure which is called the Haar measure.

2.3 L_p spaces

Definition 2.2. If (X, Ω, μ) is a measure space, then $L_p(X, \Omega, \mu)$ is the space of all measureable functions $f : X \to \mathbb{R}$ such that

$$\|f\|_p = \left[\int_X |f|^p \cdot d\mu\right]^{\frac{1}{p}} < \infty$$

For example, if $X = \mathbb{N}$, Ω is the set of all finite subsets of X, and μ is the counting measure, then $\|(x_1, x_2, \ldots, x_n, \ldots)\|_p = (\sum |x_i|^p)^{\frac{1}{p}}$. For $p = \infty$, we define

$$\|x\|_\infty = \sup_{i \in \mathbb{N}} |x_i|$$

Symmetrization is a technique that we will find useful. Loosely, the idea is that we are averaging over all the group elements.

Given a function $f : G \to \mathbb{C}$, we symmetrize it by defining $g : G \to \mathbb{C}$ as follows.

$$g(x) = \int_G f(x+a)\, d\mu(a)$$

We will use this concept in the proof of the following result.

Proposition 2.1. *If G is a locally compact abelian group, with a normalized Haar measure μ, and if $\chi_1, \chi_2 \in \hat{G}$ are two distinct characters then $\langle \chi_1, \chi_2 \rangle = 0$. i.e.,*

$$I = \int_X \chi_1(x)\overline{\chi_2(x)}\, d\mu(x) = \delta_{\chi_1,\chi_2} = \begin{cases} 0 & \chi_1 \neq \chi_2 \\ 1 & \chi_1 = \chi_2 \end{cases}$$

Proof. For any fixed $a \in G$, $I = \int_X \chi_1(x)\overline{\chi_2(x)}\, d\mu(x) = \int_X \chi_1(x+a)\overline{\chi_2(x+a)}\, d\mu(x)$. Therefore,

$$\begin{aligned} I &= \int_X \chi_1(x+a)\overline{\chi_2(x+a)}\, d\mu(x) \\ &= \int_X \chi_1(x)\chi_1(a)\overline{\chi_2(x)\chi_2(a)}\, d\mu(x) \\ &= \chi_1(a)\overline{\chi_2(a)} \int_X \chi_1(x)\overline{\chi_2(x)}\, d\mu(x) \\ &= \chi_1(a)\overline{\chi_2(a)} I \end{aligned}$$

This can only be true if either $I = 0$ or $\chi_1(a) = \chi_2(a)$. If $\chi_1 \neq \chi_2$, then there is at least one a such that $\chi_1(a) \neq \chi_2(a)$. It follows that either $\chi_1 = \chi_2$ or $I = 0$. □

By letting χ_2 be the character that is identically 1, we conclude that $\chi \in \hat{G}$ with $\chi \neq 1$ for any $\int_G \chi(x)\, d\mu(x) = 0$.

2.4 Approximation Theory

Weierstrass's theorem states that the polynomials are dense in $L_\infty[a, b] \cap C[a, b]$[1] Fejér's theorem is about approximating functions using trignometric polynomials.

Proposition 2.2. $\cos nx$ *can be expressed as a degree n polynomial in* $\cos x$.

Proof. Use the identity $\cos(u + v) + \cos(u - v) = 2 \cos u \cos v$ and induction on n. □

The polynomial $T_n(x)$ where $T_n(\cos x) = \cos(nx)$ is called n^{th} Chebyshev's polynomial. It can be seen that $T_0(s) = 1$, $T_1(s) = s$, $T_2(s) = 2s^2 - 1$ and in general $T_n(s) = 2^{n-1}s^n$ plus some lower order terms.

Theorem 2.3 (Chebyshev). *The normalized degree n polynomial $p(x) = x^n + \dots$ that approximates the function $f(x) = 0$ (on $[-1, 1]$) as well as possible in the $L_\infty[-1, 1]$ norm sense is given by $\frac{1}{2^{n-1}} T_n(x)$. i.e.,*

$$\min_{p \text{ a normalized polynomial}} \max_{-1 \leq x \leq 1} |p(x)| = \frac{1}{2^{n-1}}$$

This theorem can be proved using linear programming.

[1] This notation is intended to imply that the norm on this space is the sup-norm (clearly $C[a, b] \subseteq L_\infty[a, b]$)

2.4.1 Moment Problems

Suppose that X is a random variable. The simplest information about X are its moments. These are expressions of the form $\mu_r = \int f(x)x^r\,dx$, where f is the probability distribution function of X. A *moment problem* asks: Suppose I know all (or some of) the moments $\{\mu_r\}_{r\in\mathbb{N}}$. Do I know the distribution of X?

Theorem 2.4 (Hausdorff Moment Theorem). *If $f, g : [a,b] \to \mathbb{C}$ are two continuous functions and if for all $r = 0, 1, 2, \ldots,$ we have*

$$\int_a^b f(x)x^r\,dx = \int_a^b g(x)x^r\,dx$$

then $f = g$. Equivalently, if $h : [a,b] \to \mathbb{C}$ is a continuous function with $\int_a^b h(x)x^r\,dx = 0$ for all $r \in \mathbb{N}$, then $h \equiv 0$.

Proof. By Weierstrass's theorem, we know that for all $\epsilon > 0$, there is a polynomial P such that $\left\|\overline{h} - P\right\|_\infty < \epsilon$. If $\int_a^b h(x)x^r\,dx = 0$ for all $r \in \mathbb{N}$, then it follows that $\int_a^b h(x)Q(x)\,dx = 0$ for every polynomial $Q(x)$, and so in particular, $\int_a^b h(x)P(x)\,dx$. Therefore,

$$0 = \int_a^b h(x)P(x)\,dx = \int_a^b h(x)\overline{h(x)}\,dx + \int_a^b h(x)\left(P(x) - \overline{h(x)}\right)\,dx$$

Therefore,

$$\langle h, \overline{h}\rangle = -\int_a^b h(x)\left(P(x) - \overline{h(x)}\right)\,dx$$

Since h is continuous, it is bounded on $[a,b]$ by some constant c and so on $[a,b]$ we have $\left|h(x)\left(P(x) - \overline{h(x)}\right)\right| \le c\cdot\epsilon\cdot|b-a|$. Therefore, for any $\delta > 0$ we can pick $\epsilon > 0$ so that so that $\|h\|_2^2 \le \delta$. Hence $h \equiv 0$. $\square$

2.4.2 A little Ergodic Theory

Theorem 2.5. *Let $f : \mathbb{T} \to \mathbb{C}$ be continuous and γ be irrational. Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{r=1}^{n} f\left(e^{2\pi i r}\right) = \int_{\mathbb{T}} f(t)\,dt$$

Proof. We show that this result holds when $f(t) = e^{ist}$. Using Fejér's theorem, it will follow that the result holds for any continuous function. Now, clearly $\frac{1}{2\pi}\int_{\mathbb{T}} e^{ist}\,dt = 0$. Therefore,

$$\begin{aligned}\left|\frac{1}{n}\sum_{r=1}^{n} e^{2\pi i r s\gamma} - \frac{1}{2\pi}\int_{\mathbb{T}} e^{ist}\,dt\right| &= \left|\frac{1}{n}\sum_{r=1}^{n} e^{2\pi i r s\gamma}\right| \\ &= \left|\frac{1}{n}e^{2\pi i s\gamma}\right|\left|\frac{1-e^{2\pi i n s\gamma}}{1-e^{2\pi i s\gamma}}\right| \\ &\le \frac{2}{n\cdot(1-e^{2\pi i s\gamma})}\end{aligned}$$

Since γ is irrational, $1 - e^{2\pi i s\gamma}$ is bounded away from 0. Therefore, this quantity goes to zero, and hence the result follows. $\square$

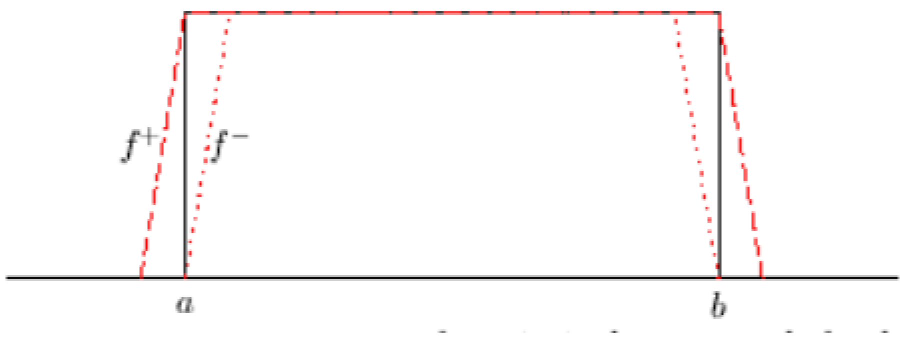

Figure 2.1: Probability of Property v. p

This result has applications in the evaluations of integrals, volume of convex bodies. Is is also used in the proof of the following result.

Theorem 2.6 (Weyl). *Let γ be an irrational number. For $x \in \mathbb{R}$, we denote by $\langle x \rangle = x - [x]$ the fractional part of x. For any $0 < a < b < 1$, we have*

$$\lim_{n \to \infty} \frac{|\{1 \leq r \leq n : a \leq \langle r\gamma \rangle < b\}|}{n} = b - a$$

Proof. We would like to use Theorem 2.5 with the function $f = 1_{[a,b]}$. However, this function is not continuous. To get around this, we define functions $f^+ \geq 1_{[a,b]} \geq f^-$ as shown in the following diagram.

f^+ and f^- are continuous functions approximating f. We let let them approach f and pass to the limit. □

This is related to a more general ergodic theorem by Birkhoff.

Theorem 2.7 (Birkhoff, 1931). *Let $(\Omega, \mathcal{F}, p)$ be a probability measure and $T : \Omega \to \Omega$ be a measure preserving transformation. Let $X \in L_1(\Omega, \mathcal{F}, p)$ be a random variable. Then*

$$\frac{1}{n} \sum_{k=1}^{n} X \circ T^k \to E[X; \mathcal{I}]$$

Where $\mathcal{I}$ is the σ-field of T-invariant sets.

2.5 Some Convergence Theorems

We seek conditions under which $S_n(f,t) \to f(t)$ (preferably uniformly). Some history:

- DuBois Raymond gave an example of a continuous function such that $\limsup S_n(f,0) = \infty$.
- Kolmogorov [1] found a Lebesgue measureable function $f : \mathbb{T} \to \mathbb{R}$ such that for all t, $\limsup S_n(f,t) = \infty$.

- Carleson [2] showed that if $f : \mathbb{T} \to \mathbb{C}$ is a continuous function (even Riemann integrable), then $S_n(f,t) \to f(t)$ almost everywhere.

- Kahane and Katznelson [3] showed that for every $E \subseteq \mathbb{T}$ with $\mu(E) = 0$, there exists a continuous function $f : \mathbb{T} \to \mathbb{C}$ such that $S_n(f,t) \not\to f(t)$ if and only if $t \in E$.

Definition 2.3. $\ell_p = L_p(\mathbb{N}, \text{Finite sets}, \text{ counting measure}). = \{\boldsymbol{x} | (x_0, \dots)|^p < \infty\}$.

Theorem 2.8. *Let $f : \mathbb{T} \to \mathbb{C}$ be continuous and suppose that $\sum_{r \in \mathbb{Z}} |\hat{f}(r)| < \infty$ (so $\hat{f} \in \ell_1$). Then $S_n(f,t) \to f$ uniformly on $\mathbb{T}$.*

Proof. See lecture 3, theorem 3.1. □

2.6 The L_2 theory

The fact that $e(t) = e^{ist}$ is an orthonormal family of functions allows to develop a very satisfactory theory. Given a function f, the best coefficients $\lambda_1, \lambda_2, \dots, \lambda_n$ so that $\|f - \sum_{i=1}^{n} \lambda_j e_j\|_2$ is minimized is given by $\lambda_j = \langle f, e_j \rangle$. This answer applies just as well in any inner product normed space (Hilbert space) whenever $\{e_j\}$ forms an orthonormal system.

Theorem 2.9 (Bessel's Inequality). *For every $\lambda_1, \lambda_2, \dots, \lambda_n$,*

$$\left\| f - \sum_{i=1}^{n} \lambda_i e_i \right\|^2 \geq \|f\|^2 - \sum_{i=1}^{n} \langle f, e_i \rangle^2$$

with equality when $\lambda_i = \langle f, e_i \rangle$

Proof. We offer a proof here for the real case, in the next lecture the complex case will be done as well.

$$\begin{aligned} \left\| f - \sum_{i=1}^{n} \lambda_i e_i \right\|^2 &= \left\| (f - \sum_{i=1}^{n} \langle f, e_i \rangle e_i) + (\sum_{i=1}^{n} \langle f, e_i \rangle e_i - \sum_{i=1}^{n} \lambda_i e_i) \right\|^2 \\ &= \left\| (f - \sum_{i=1}^{n} \langle f, e_i \rangle e_i) \right\|^2 + \left\| (\sum_{i=1}^{n} \langle f, e_i \rangle e_i - \sum_{i=1}^{n} \lambda_i e_i) \right\|^2 + \text{ cross terms} \end{aligned}$$

$$\text{cross terms } = 2\langle f - \sum_{i=1}^{n} \langle f, e_i \rangle e_i, \sum_{i=1}^{n} \langle f, e_i \rangle e_i - \sum_{i=1}^{n} \lambda_i e_i \rangle$$

Observe that the terms in the cross terms are orthogonal to one another since $\forall i \langle f - \langle f, e_i \rangle e_i, e_i \rangle = 0$. We write

$$2 \sum \langle f, e_i \rangle \langle f - \sum_{j=1}^{n} \langle f, e_j \rangle e_j, e_i \rangle - \sum_{i}^{n} \lambda_i \langle f - \sum_{j=1}^{n} \langle f, e_j \rangle e_i, e_i \rangle$$

Observe that each innter product term is 0. Since if $i = j$, then we apply $\forall i \langle f - \langle f, e_i \rangle e_i, e_i \rangle = 0$. If $i \neq j$, then they are orthogonal basis vectors.

We want to make this as small as possible and have only control over the λ_is. Since this term is squared and therefore non-negative, the sum is minimized when we set $\forall i\ \lambda_i = \langle f, e_i \rangle$. With this choice,

$$\begin{aligned}\left\| f - \sum_{i=1}^{n} \lambda_i e_i \right\|^2 &= \langle f - \sum_{i=1}^{n} \lambda_i e_i, f - \sum_{i=1}^{n} \lambda_i e_i \rangle \\ &= \langle f, f \rangle - 2 \sum_{i=1}^{n} \lambda_i \langle f, e_i \rangle + \sum_{i=1}^{n} \lambda_i^2 \\ &= \|f\|^2 - \sum_{i=1}^{n} \langle f, e_i \rangle^2\end{aligned}$$

where the last inequality is obtained by setting $\lambda_i = \langle f, e_i \rangle$. □

References

[1] A. N. Kolmogorov, *Une série de Fourier-Lebesgue divergente partout*, CRAS Paris, 183, pp. 1327-1328, 1926.

[2] L. Carleson, *Convergence and growth of partial sums of Fourier series*, Acta Math. 116, pp. 135-157, 1964.

[3] J-P Kahane and Y. Katznelson, *Sur les ensembles de divergence des séries trignométriques*, Studia Mathematica, 26 pp. 305-306, 1966