

# STAT 7200

## Introduction to Advanced Probability

### Lecture 21

Taylor R. Brown

## 1 (Levy's) Continuity Theorem

# (Levy's) Continuity Theorem

## Theorem 1 (The Continuity Theorem (11.1.14))

*Let  $\mu, \mu_1, \mu_2, \dots$  be probability measures with characteristic functions  $\phi, \phi_1, \phi_2, \dots$ . Then  $\mu_n$  converges weakly to  $\mu$  if and only if  $\phi_n(t) \rightarrow \phi(t)$  for all  $t \in \mathbb{R}$ . That is, the weak convergence is equivalent to the pointwise convergence of characteristic functions.*

- **Proof:** (1) *Weak convergence implies pointwise convergence of characteristic functions:*
  - Since  $\cos(x)$  and  $\sin(x)$  are both bounded and continuous functions, then for any  $t \in \mathbb{R}$  :

$$\begin{aligned}\phi_n(t) &= \int \cos(tx) \mu_n(dx) + i \int \sin(tx) \mu_n(dx) \\ &\rightarrow \int \cos(tx) \mu(dx) + i \int \sin(tx) \mu(dx) = \phi(t),\end{aligned}$$

by the definition of weak convergence.

# Continuity Theorem: Pointwise Convergence of Characteristic Function Implies Weak Convergence

- **Proof:** (2) On the other hand, if we have  $\phi_n(t) \rightarrow \phi(t)$  for all  $t \in \mathbb{R}$ , we do not even know if the limit of  $\{\mu_n\}$  exists or not.
- We will need several theorems, lemmas and corollaries to show that this is indeed true. Many of these will need their own results to prove them.

# Fourier Inversion Theorem

## Theorem 2 (Fourier Inversion Theorem (11.1.1))

Let  $\mu$  be a Borel probability measure on  $\mathbb{R}$  with characteristic function  $\phi(t) = \int e^{itx} \mu(dx)$ . Then for  $a < b$  and  $\mu(\{a\}) = \mu(\{b\}) = 0$ :

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

- **Remark 1:** We'll prove this with two Lemmas.
- **Remark 2:** The number of intervals  $[a, b]$  with  $\mu(\{a\}) \neq 0$  or  $\mu(\{b\}) \neq 0$  is at most countable because the set  $\{x : \mu(\{x\}) > 0\}$  is at most countable. That's from the previous lecture.

# First Lemma to prove Fourier Inversion Theorem

## Theorem 3 (Lemma 11.1.2)

For  $T \geq 0$  and  $a < b$

$$\int_{\mathbb{R}} \int_{-T}^T \left| \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \right| dt \mu(dx) \leq 2T(b-a) < \infty.$$

## First Lemma to prove Fourier Inversion Theorem

$$\begin{aligned}\left| \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \right| &= \left| \frac{e^{-ita} - e^{-itb}}{it} \right| |e^{itx}| \\ &= \left| \int_a^b e^{itr} dr \right| \\ &\leq \int_a^b |e^{itr}| dr \\ &= b - a\end{aligned}$$

So

$$\int_{\mathbb{R}} \int_{-T}^T \left| \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \right| dt \mu(dx) \leq \int_{\mathbb{R}} \int_{-T}^T (b - a) dt \mu(dx) = 2T(b - a)$$

## Second Lemma to prove Fourier Inversion Theorem

### Theorem 4 (Lemma 11.1.3)

For  $T \geq 0$  and  $\theta \in \mathbb{R}$

$$\lim_{T \rightarrow \infty} \int_{-T}^T \frac{\sin(\theta t)}{t} dt = \pi \operatorname{sign}(\theta)$$

where  $\operatorname{sign}(\theta)$  is either 1,  $-1$  or 0 depending on whether  $\theta$  is positive, negative or 0, respectively.

We're omitting the proof because it's elementary integration, but it's fun and involves a lot of cool stuff (e.g. the sinc function, integration by parts, u-substitution, different trigonometric properties, etc.) so you should try it.



# Fourier Inversion Theorem: Proof

WTS:

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

$$\begin{aligned} & \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt \\ &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \left( \int_{\mathbb{R}} e^{itx} \mu(dx) \right) dt \quad (\text{Fubini and first Lemma}) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{-T}^T \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt \mu(dx) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{-T}^T \frac{i \sin(t(x-a)) - i \sin(t(x-b))}{it} dt \mu(dx) \end{aligned}$$

## Fourier Inversion Theorem: Proof (continued)

Taking  $T \rightarrow \infty$ :

$$\begin{aligned} & \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{\mathbb{R}} \int_{-T}^T \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} dt \mu(dx) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} dt \mu(dx) \quad (\text{DCT}) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \pi [\text{sign}(x-a) - \text{sign}(x-b)] \mu(dx) \quad (\text{Lemma 2}) \\ &= \mu((a, b)) + \frac{1}{2} \mu(\{a\}) + \frac{1}{2} \mu(\{b\}) \\ &= \mu([a, b]) \end{aligned}$$

# Fourier Uniqueness Theorem: Characteristic Function Determines Distribution

## Theorem 5 (Fourier Uniqueness Theorem)

*Let  $X, Y$  be random variables. Then  $\phi_X(t) = \phi_Y(t)$  if and only if  $\mathcal{L}(X) = \mathcal{L}(Y)$ .*

- **Proof:** The “if” part is trivial. For the “only if” part, to show  $\mathcal{L}(X) = \mathcal{L}(Y)$ , we only need to show  $P(X \in I) = P(Y \in I)$  for all intervals in  $\mathbb{R}$  (uniqueness of extension).
  - By Fourier Inversion theorem,  $P(X \in [a, b]) = P(Y \in [a, b])$  whenever  $a < b$  and  $P(X = a) = P(X = b) = P(Y = a) = P(Y = b) = 0$ .
  - For any interval  $I$ , we can always find a sequence of closed intervals  $\{[a_i, b_i]\}$  that satisfy the above conditions, and  $[a_i, b_i] \rightarrow I$ . Thus, we can apply the continuity of probability to show that  $P(X \in I) = P(Y \in I)$ .

# Helly Selection Principle

## Theorem 6 (Helly Selection Principle)

*Let  $\{F_n\}$  be a sequence of cdfs, each corresponding with a measure  $\mu_n$ . Then there exists a subsequence  $F_{n_k}$ , and a non-decreasing, right-continuous  $0 \leq F \leq 1$ , such that  $F_{n_k}(x) \rightarrow F(x)$  for each  $x \in \mathbb{R}$  that is a continuity point of  $F$ .*

The proof will remind you of the uncountability proofs.

Also,  $F$  is not necessarily a cdf.

## Helly Selection Principle: proof

List out rationals  $Q = \{q_1, q_2, \dots\}$ . Note that  $0 \leq F_n(q_1) \leq 1$  for all  $n$ . By Bolzano-Weirstrass, there exists a subsequence  $I_k^{(1)}$  such that  $\lim_k F_{I_k^{(1)}}(q_1)$  exists. Then there exists a subsequence of that subsequence, call it  $I_k^{(2)}$ , such that  $\lim_k F_{I_k^{(2)}}(q_2)$  exists. Because it is a subsequence of the first one, we also have that  $\lim_k F_{I_k^{(2)}}(q_1)$  exists. We can do this for each  $m \in \mathbb{N}$ . For each  $\{I_k^{(m)}\}$ , we have  $\lim_k F_{I_k^{(m)}}(q_j)$  exists for all  $0 < j \leq m$ .

Now define  $n_k = I_k^{(k)}$ . These are the diagonals. However, note that all these subsequences are nested, so for any  $k$ ,

- $\{n_k, n_{k+1}, \dots\} \subseteq \{I_k^{(k)}, I_{k+1}^{(k)}, \dots\}$
- $\{n_{k+1}, n_{k+2}, \dots\} \subseteq \{I_{k+1}^{(k+1)}, I_{k+2}^{(k+1)}, \dots\}$ , etc.

These ensure that  $\lim_k F_{n_k}(q) := G(q)$  exists for each  $q \in Q$ .  $G$  is also clearly non-decreasing as well, just like each  $F_{n,k}$ .

## Helly Selection Principle: proof

For each rational  $q$ ,  $F_{n_k}(q) \rightarrow G(q)$ .  $G$  is defined on the rationals, only. Now we define

$$F(x) = \inf\{G(q) : q > x, q \in \mathbb{Q}\}$$

which is defined on the reals. It has a few properties that we'll need:

- $F$  is non-decreasing
- $0 \leq F \leq 1$
- $F$  is right-continuous, and
- $F(q) > G(q)$  for all  $q \in \mathbb{Q}$

Next, we'll show that, for any continuity point of  $F$ , call it  $x \in \mathbb{R}$ , we have  $F_{n_k}(x) \rightarrow F(x)$  as  $k \rightarrow \infty$ .

Pick any  $\varepsilon > 0$ , then pick  $r, u, s \in \mathbb{Q}$  such that  $r < u < x < s$  and  $F(s) - F(r) < \varepsilon$ .

$$\begin{aligned} F(x) - \varepsilon &\leq F(r) \\ &= \inf\{G(q) : q > r\} \\ &= \inf\{\lim_k F_{n_k}(q) : q > r\} \end{aligned}$$

# Helly Selection Principle: proof

$$\begin{aligned} F(x) - \varepsilon &\leq \inf_{q > r} \liminf_k F_{n_k}(q) \\ &\leq \liminf_k F_{n_k}(u) && (u > r) \\ &\leq \liminf_k F_{n_k}(x) && (x > u) \\ &\leq \limsup_k F_{n_k}(x) \\ &\leq \limsup_k F_{n_k}(s) && (s > x) \\ &= G(s) \\ &\leq F(s) \\ &\leq F(x) + \varepsilon \end{aligned}$$

QED

# Tightness of Measure: Three More Results

## Theorem 7 (11.1.10)

*If  $\{\mu_n\}$  is a tight sequence of prob. measures, then there exists a subsequence  $\{\mu_{n_k}\}$  and a probability measure  $\mu$  such that  $\mu_{n_k}$  converges weakly to  $\mu$ .*

## Theorem 8 (Corollary 11.1.11)

*Let  $\{\mu_n\}$  be a tight sequence of prob. measures on  $\mathbb{R}$ . Also suppose that, whenever  $\mu_{n_k} \Rightarrow \nu$ , then  $\nu$  is always equal to  $\mu$ . Then  $\mu_n \Rightarrow \mu$ .*

## Lemma 9 (11.1.13)

*Let  $\{\mu_n\}$  be a sequence of probability measures on  $\mathbb{R}$ , and  $\{\phi_n(t)\}$  be the characteristic functions. If there is a function  $g$  that is continuous at 0, and  $\phi_n(t) \rightarrow g(t)$  for all  $|t| < t_0$  ( $t_0 > 0$ ), then  $\{\mu_n\}$  is tight.*

Remember our goal is to prove the other direction of Levy's continuity theorem. Let's prove the third one first.



# Tightness and Characteristic Functions

## Lemma 10 (11.1.13)

Let  $\{\mu_n\}$  be a sequence of probability measures on  $\mathbb{R}$ , and  $\{\phi_n(t)\}$  be the characteristic functions. If there is a function  $g$  that is continuous at 0, and  $\phi_n(t) \rightarrow g(t)$  for all  $|t| < t_0$  ( $t_0 > 0$ ), then  $\{\mu_n\}$  is tight.

• **Proof:** Let  $y > 0$

$$\begin{aligned} \frac{1}{y} \int_{-y}^y [1 - \phi_n(t)] dt &= \int_{-\infty}^{\infty} \left[ \frac{1}{y} \int_{-y}^y (1 - e^{itx}) dt \right] \mu_n(dx) \\ &= 2 \int_{-\infty}^{\infty} \left( 1 - \frac{\sin yx}{yx} \right) \mu_n(dx) \\ &\geq \int_{|x| > 2/y} \left( 1 - \frac{1}{|yx|} \right) \mu_n(dx) \\ &\geq \mu_n \left[ \left\{ x : |x| \geq \frac{2}{y} \right\} \right] \end{aligned}$$

## Tightness and Characteristic Functions: continued

- **Proof continued:** The previous discussion shows that
$$\mu_n[\{x : |x| \geq \frac{2}{y}\}] \leq \frac{1}{y} \int_{-y}^y [1 - \phi_n(t)] dt.$$
- Now since  $g(t)$  is continuous at 0, and  $g(0) = \lim_n \phi_n(0) = 1$ , then for any  $\varepsilon > 0$ , we can always find  $y_0 \in (0, t_0)$  such that:

$$|1 - g(t)| < \varepsilon/4$$

whenever  $|t| < y_0$ . Then

$$\begin{aligned} \left| \frac{1}{y_0} \int_{-y_0}^{y_0} [1 - g(t)] dt \right| &\leq \frac{1}{y_0} \int_{-y_0}^{y_0} |1 - g(t)| dt \\ &\leq \frac{1}{y_0} \int_{-y_0}^{y_0} \varepsilon/4 dt = \varepsilon/2 \end{aligned}$$

# Tightness and Characteristic Functions: continued

- **Proof continued:** The previous discussion shows that

①  $\mu_n[\{x : |x| \geq \frac{2}{y}\}] \leq \frac{1}{y} \int_{-y}^y [1 - \phi_n(t)] dt$  and

②  $\left| \frac{1}{y_0} \int_{-y_0}^{y_0} [1 - g(t)] dt \right| \leq \varepsilon/2$

- On the other hand, as  $\phi_n(t) \rightarrow g(t)$  for  $|t| < t_0$ , and  $|\phi_n(t)| = 1$ , we can apply the dominated convergence theorem:

$$\left| \int_{-y_0}^{y_0} [1 - \phi_n(t)] dt - \int_{-y_0}^{y_0} [1 - g(t)] dt \right| \leq \varepsilon/2$$

for  $n > N$ .

- Then for all  $n > N$ ,

$$\mu_n[\{x : |x| \geq \frac{2}{y_0}\}] \leq \frac{1}{y_0} \int_{-y_0}^{y_0} [1 - \phi_n(t)] dt \leq \varepsilon$$

- It then follows that  $\{\mu_n\}$  must be tight.