Math 525: Lecture 14

February 22, 2018

1 Tightness

Last lecture, we discussed convergence in distribution, culminating in Helly's theorem:

Proposition 1.1 (Helly's theorem). *Let $(F_n)_n$ be a sequence of distribution functions. Then, there exists a subsequence $(n_k)_k$ and a right continuous nondecreasing function F such that*

$$F_{n_k}(x) \to F(x) \text{ for all continuity points } x \text{ of } F.$$

The issue with the above is that F need not be a distribution function:

Example 1.2.

1. Let $X_n = n$. Then, $F_n(x) = I_{[n,\infty)}(x)$ and $F_n(x) \to 0$ for all x.
2. Let $X_n = -n$. Then, $F_n(x) = I_{[-n,\infty)}(x)$ and $F_n(x) \to 1$ for all x.

Neither $F = 0$ nor $F = 1$ are distribution functions.

A criteria that ensures the limiting function is indeed a distribution function is tightness:

Definition 1.3. Let $\{F_\alpha\}_\alpha$ be a family of distribution functions. We say $\{F_\alpha\}_\alpha$ is *tight* if for every $\epsilon > 0$, there exists r sufficiently large such that

$$F_\alpha(r) - F_\alpha(-r) \geq 1 - \epsilon$$

for all α.

Proposition 1.4. *Suppose that $(F_n)_n$ is a tight sequence of distribution functions. Then, there exists a subsequence $(n_k)_k$ and a distribution function F such that $F_{n_k} \Rightarrow F$.*

In other words, the space of distribution functions is *sequentially compact*.

Proof. By Helly's theorem, we can find a subsequence $(n_k)_k$ and a right continuous nondecreasing function F such that

$$F_{n_k}(x) \to F(x) \text{ for all continuity points } x \text{ of } F.$$

By tightness, we can find r such that

$$F_n(-r) + (1 - F_n(r)) \leq \epsilon.$$

Since $F_n(-r)$ and $1 - F_n(r)$ are both nonnegative, this implies

$$F_n(-r) \leq \epsilon \qquad \text{and} \qquad 1 - F_n(r) \leq \epsilon.$$

Now, choose $x_\epsilon > r$ so that both x_ϵ and $-x_\epsilon$ are continuity points of F. Then,

$$F(-x_\epsilon) = \lim_k F_{n_k}(-x_\epsilon) \leq \epsilon$$

and

$$1 - F(x_\epsilon) = \lim_k \{1 - F_{n_k}(x_\epsilon)\} \leq \epsilon.$$

Since ϵ was arbitrary, this implies

$$\lim_{x \to -\infty} F(x) = 0 \qquad \text{and} \qquad \lim_{x \to \infty} F(x) = 1,$$

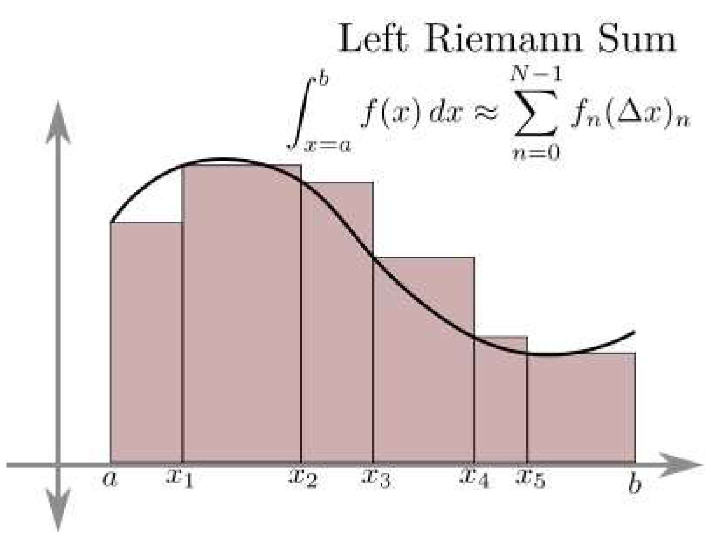
as desired. □

2 Integration

We will work a lot with integrals today, so let's digress and briefly talk about integration. Suppose $f \colon \mathbb{R} \to \mathbb{R}$ is a continuous function. Then, the integral

$$\int_a^b f(t)dt$$

can be interpreted as the limit of Riemann sums.



What about when f is not a "nice" function? Let $Y \sim U[0,1]$ and define

$$\int_a^b f(t)dt \equiv (b-a)\,\mathbb{E}\left[f(Y)\right].$$

The above extends the theory of integration to Borel measurable functions f (recall that if f is Borel measurable, $f \circ Y$ is a random variable). You can check that the definition of the integral above satisfies all the usual conditions (e.g., linearity) and agrees with the Riemann integral when f is "nice".

Remark 2.1. You may have seen the Lebesgue integral. The above is not quite the Lebesgue integral, since it is only defined for Borel measurable functions f.

The above tells us that we can apply things like the Monotone Convergence Theorem, Fatou's Lemma, and the Dominated Convergence Theorem to regular integrals by treating them like expectations!

3 Lévy's continuity theorem

Our next goal is to establish Paul Lévy's continuity theorem, which (roughly speaking) estabishes that a sequence of random variables converges in distribution if and only if their characteristic functions converge pointwise.

3.1 Forward direction

We have already done all the hard work to prove the forward direction.

Proposition 3.1. *If $X_n \xrightarrow{\mathcal{D}} X$ (i.e., $F_n \Rightarrow F$) then $\mathbb{E}[e^{itX_n}] \equiv \phi_n(t) \to \phi(t) \equiv \mathbb{E}[e^{itX}]$ for all t.*

Proof. Remember that convergence in distribution was equivalent to

$$\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X)\right]$$

for all bounded and continuous f. Let t be arbitrary and take $f(x) = e^{itx}$ so that

$$\phi_n(t) = \mathbb{E}\left[e^{itX_n}\right] \to \mathbb{E}\left[e^{itX}\right] = \phi(t). \qquad \square$$

3.2 Reverse direction

The remainder of this section is dedicated to proving the reverse direction. Before we do so, let X be a random variable and ϕ its characteristic function. Then, for any $T > 0$,

$$\begin{aligned}\frac{1}{2T}\int_{-T}^{T}\phi(t)dt &= \frac{1}{2T}\int_{-T}^{T}\mathbb{E}\left[e^{itX}\right]dt \\ &= \frac{1}{2T}\mathbb{E}\left[\int_{-T}^{T}e^{itX}dt\right] = \frac{1}{2T}\mathbb{E}\left[\frac{2\sin(TX)}{X}\right] = \mathbb{E}\left[\frac{\sin(TX)}{TX}\right]\end{aligned}$$

where we have used the Fubini-Tonelli theorem to move the integration into the expectation (take this as fact if you have yet to encounter Fubini-Tonelli). Now, let $A > 0$ and note that if $|x| > 2A$, then

$$\left|\frac{\sin(Tx)}{Tx}\right| \leq \frac{1}{2TA}.$$

Let $B = \{-2A < X \leq 2A\}$. Then,

$$\begin{aligned}
\left|\mathbb{E}\left[\frac{\sin(TX)}{TX}\right]\right| &\leq \mathbb{E}\left[\left|\frac{\sin(TX)}{TX}\right| I_B + \left|\frac{\sin(TX)}{TX}\right| I_{B^c}\right] \\
&\leq \mathbb{E}\left[I_B + \frac{1}{2TA} I_{B^c}\right] \\
&= (F(2A) - F(-2A)) + \frac{1 - (F(2A) - F(-2A))}{2TA} \\
&= \left(1 - \frac{1}{2TA}\right)(F(2A) - F(-2A)) + \frac{1}{2TA}.
\end{aligned}$$

Now, take $T = 1/A$ to get

$$\frac{A}{2}\left|\int_{-1/A}^{1/A} \phi(t)dt\right| \leq \frac{1}{2}(F(2A) - F(-2A)) + \frac{1}{2}.$$

Some algebra reveals

$$A\left|\int_{-1/A}^{1/A} \phi(t)dt\right| - 1 \leq F(2A) - F(-2A). \tag{1}$$

In particular, (1) gives a criterion for tightness in terms of characteristic functions:

Proposition 3.2. *Let $(X_n)_n$ be a sequence of random variables with distribution and characteristic functions F_n and ϕ_n. The sequence $(F_n)_n$ is tight if for all $\epsilon > 0$, there exists $\delta > 0$ such that*

$$\frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi_n(t)dt\right| - 1 \geq 1 - \epsilon.$$

Proof. Take $A = 1/\delta$ in (1). □

We are now ready to prove the reverse direction.

Proposition 3.3 (Lévy's continuity theorem)**.** *Let $(F_n)_n$ be a sequence of distribution functions and ϕ_n be the characteristic function of F_n. Suppose $\phi_n \to \phi$ pointwise for some function ϕ which is continuous at the origin ($t = 0$). Then, there exists a distribution function F such that $F_n \Rightarrow F$ and ϕ is the characteristic function of F.*

Some notes:

- When we say "ϕ is the characteristic function of F", we mean that a random variable X associated to F (e.g., take $X = F^{-1}(Y)$ where $Y \sim U[0,1]$) has characteristic function ϕ.

- That ϕ is a characteristic function is one of the results of the theorem (the limit of characteristic functions need not be a characteristic function in general).

Proof. Since ϕ is continuous at zero, we can choose δ small enough so that

$$|\phi(t) - 1| < \epsilon/4 \qquad \text{whenever } |t| < \delta.$$

Now, for $Y \sim U[-\delta, \delta]$,

$$\frac{1}{\delta}\int_{-\delta}^{\delta} \phi_n(t)dt \equiv 2\mathbb{E}\left[\phi_n(Y)\right] \to 2\mathbb{E}\left[\phi(Y)\right] \equiv \frac{1}{\delta}\int_{-\delta}^{\delta} \phi(t)dt$$

by the DCT. Therefore, we can choose N large enough such that for all $n \geq N$,

$$\frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi_n(t) - \phi(t)dt\right| < \frac{\epsilon}{2}.$$

Next, note that

$$\begin{aligned}
\frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi(t)dt\right| &= \frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi(t) - \phi_n(t) + \phi_n(t)dt\right| \\
&\leq \frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi(t) - \phi_n(t)dt\right| + \frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi_n(t)dt\right| \\
&< \frac{\epsilon}{2} + \frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi_n(t)dt\right|.
\end{aligned}$$

Recalling that $|\phi(t) - 1| < \epsilon/2$ for all $t \in [-\delta, \delta]$, it follows that

$$\frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi(t)dt\right| \equiv 2\left|\mathbb{E}\left[\phi(Y)\right]\right| > 2\left(1 - \frac{\epsilon}{4}\right) = 2 - \frac{\epsilon}{2}.$$

Therefore,

$$2 - \frac{\epsilon}{2} < \frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi(t)dt\right| < \frac{\epsilon}{2} + \frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi_n(t)dt\right|.$$

Simplying,

$$2 - \epsilon < \frac{1}{\delta}\left|\int_{-\delta}^{\delta} \phi_n(t)dt\right|.$$

By Proposition 3.2, the sequence $(F_n)_n$ is tight. By Helly's theorem, there exists a subsequence $(F_{n_k})_k$ and a distribution function F such that $F_{n_k} \Rightarrow F$. Let $\tilde{\phi}$ be the characteristic function of F. Then, by Proposition 3.1, $\phi_{n_k} \to \tilde{\phi}$ pointwise. But $\phi_n \to \phi$, and hence $\phi = \tilde{\phi}$.

It can be shown that $F_n \Rightarrow F$, but the proof requires showing that convergence in distribution is metrizable, which is out of the scope of this course. □

We can finally return to a claim we made a long time ago about the relationship between Poisson and Binomial random variables:

Example 3.4. The characteristic function of Poisson(λ) is

$$\phi_{\text{Poisson}(\lambda)}(t) = \exp\left(\lambda\left(e^{it} - 1\right)\right).$$

The characteristic function of $B(n, p_n)$ is

$$\phi_{B(n,p_n)}(t) = \left((1 - p_n) + p_n e^{it}\right)^n.$$

Suppose $np_n \to \lambda$ as $n \to \infty$. Then,

$$\left((1 - p_n) + p_n e^{it}\right)^n = \left(1 - np_n\left(e^{it} - 1\right)\frac{1}{n}\right)^n \to \exp\left(\lambda\left(e^{it} - 1\right)\right).$$