Math 525: Lecture 15

March 6, 2018

Today, we finally prove the central limit theorem (CLT). Recall that for a "sufficiently nice" sequence of i.i.d. random variables $(X_n)_n$, the law of large numbers told us

$$\frac{S_n}{n} \equiv \frac{X_1 + \cdots + X_n}{n} \to \mathbb{E}X_1 \text{ a.s.}$$

The CLT will tell us about the distribution of S_n/n. Namely,

$$\sqrt{n}\left(\frac{S_n}{n} - \mathbb{E}X_1\right) \xrightarrow{\mathcal{D}} Y \tag{1}$$

where $Y \sim \mathcal{N}(0, \mathrm{Var}(X_1))$. Before we give the CLT, let's review normal random variables.

1 Normal random variables

Definition 1.1. We say X is a *normal random variable* with mean μ and variance σ^2 if its probability density is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

In this case, we write $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proposition 1.2. *The characteristic function of $X \sim \mathcal{N}(0,1)$ random variable is*

$$\phi_X(t) = e^{-t^2/2}.$$

Proof. Note that

$$\begin{aligned}
\phi_X(t) = \mathbb{E}\left[e^{itX}\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{itx} dx \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} e^{-\frac{x^2}{2}} e^{itx} dx + \int_{0}^{\infty} e^{-\frac{x^2}{2}} e^{itx} dx\right) \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} e^{-\frac{x^2}{2}} e^{itx} dx + \int_{0}^{\infty} e^{-\frac{x^2}{2}} e^{-itx} dx\right) \\
&= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{x^2}{2}} \cos(tx) dx.
\end{aligned}$$

Now, take the derivative with respect to t to get

$$\phi_X'(t) = -\frac{2}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} x \sin(tx) dx.$$

Integrate by parts to get

$$\begin{aligned}
\phi_X'(t) &= \frac{2}{\sqrt{2\pi}} \left(e^{-\frac{x^2}{2}} \sin(tx) \mid_0^\infty -t \int_0^\infty e^{-\frac{x^2}{2}} \cos(tx) dx \right) \\
&= -\frac{2t}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} \cos(tx) dx \\
&= -t\phi_X(t).
\end{aligned}$$

Note that

$$\phi'(t) = -t\phi(t)$$

is an ordinary differential equation with

$$\phi(t) = C_0 \exp\left(-\frac{t^2}{2}\right)$$

where C_0 is some constant. Since ϕ is a characteristic function, we must have $\phi(0) = 1$. This implies that $C_0 = 1$. $\square$

2 Classical CLT

Just like with the law of large numbers, there are various versions of the CLT. Here is the first version we encounter:

Proposition 2.1. *Let $(X_j)_j$ be a sequence of i.i.d. random variables. Let $\mu = \mathbb{E}[X_1]$, $\sigma^2 = \mathrm{Var}(X_1)$, and $S_n = X_1 + \cdots + X_n$. Then,*

$$\frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mu \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

(note that this is identical to (1) if we multiply both sides by σ).

In establishing Levy's continuity theorem, we have actually done all the hard work already to prove this fact.

Proof. First, note that

$$\frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mu \right) = \frac{(X_1 - \mu) + \cdots + (X_n - \mu)}{\sqrt{n}\sigma} = \frac{Y_1 + \cdots + Y_n}{\sqrt{n}}.$$

where

$$Y_n = \frac{X_n - \mu}{\sigma}.$$

Note that

$$\mathbb{E}Y_n = \frac{\mathbb{E}X_n - \mu}{\sigma} = 0$$
$$\operatorname{Var}(Y_n) = \operatorname{Var}\left(\frac{X_n - \mu}{\sigma}\right) = 1.$$

For brevity, let $S'_n = Y_1 + \cdots + Y_n$.

The characteristic function of Y_n is

$$\begin{aligned}
\phi(t) &= \mathbb{E}[e^{itY_n}] \\
&= \mathbb{E}\left[\sum_{k\geq 0} \frac{(it)^k}{k!} Y_n^k\right] \\
&= \mathbb{E}\left[Y^0\right] + (it)\,\mathbb{E}\left[Y^1\right] + \frac{(it)^2}{2}\mathbb{E}\left[Y^2\right] + \cdots \\
&= 1 - \frac{t^2}{2} + \cdots
\end{aligned}$$

More precisely,

$$\phi(t) = 1 - \frac{t^2}{2} + h(t^2)$$

where h denotes a function that satisfying $h(cx)/x \to 0$ as $x \to 0$ (corresponding to the higher order terms in the Taylor expansion). Now, let

$$S'_n = \frac{Y_1 + \cdots + Y_n}{\sqrt{n}}.$$

Note that

$$\begin{aligned}
\phi_{S'_n}(t) &= \mathbb{E}\left[\exp\left(it\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}\right)\right] \\
&= \mathbb{E}\left[\exp\left(it\frac{Y_1}{\sqrt{n}}\right)\cdots\exp\left(it\frac{Y_n}{\sqrt{n}}\right)\right] \\
&= \mathbb{E}\left[\exp\left(it\frac{Y_1}{\sqrt{n}}\right)\right]\cdots\mathbb{E}\left[\exp\left(it\frac{Y_n}{\sqrt{n}}\right)\right] \\
&= \phi\left(\frac{t}{\sqrt{n}}\right)\cdots\phi\left(\frac{t}{\sqrt{n}}\right) \\
&= \left(\phi\left(\frac{t}{\sqrt{n}}\right)\right)^n \\
&= \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n. \\
&= \left(1 + \frac{z_n}{n}\right)^n
\end{aligned}$$

where we have defined

$$z_n = -\frac{t^2}{2} + nh\left(\frac{t^2}{n}\right).$$

Let $c = t^2$ and $x = 1/n$. Then,

$$nh\left(\frac{t^2}{n}\right) = \frac{h\left(cx\right)}{x} \to 0 \qquad \text{as} \qquad x \to 0.$$

Therefore,

$$\phi_{S_n'}(t) \to e^{t^2/2}.$$

By Lévy's continuity theorem, S_n' converges to a standard normal random variable. $\square$

Exercise 2.2. n numbers are rounded to the nearest integer and then summed. Suppose that the individual roundoff errors are uniformly distributed over $[-0.5, 0.5]$. What is the probability that the roundoff error exceeds the exact sum by more than x?

Hint: let Y_j be the j-th number and $[Y_j]$ be the result from rounding. Then,

$$X_j = Y_j - [Y_j] \sim U[-0.5, 0.5].$$

Therefore,

$$\mathbb{E}X_j = 0 \qquad \text{and} \qquad \sigma^2 = \mathrm{Var}(X_j) = \mathbb{E}X_j^2 = \int_{-1/2}^{1/2} x^2 dx = \frac{1}{12}.$$

Let $S_n = X_1 + \cdots + X_n$. Then,

$$\begin{aligned}
\mathbb{P}\left\{|S_n| > x\right\} &= 1 - \mathbb{P}\left\{|S_n| \leq x\right\} \\
&= 1 - \mathbb{P}\left\{\frac{|S_n|}{\sqrt{n}\sigma} \leq \frac{x}{\sqrt{n}\sigma}\right\} \\
&= 1 - \mathbb{P}\left\{-\frac{x}{\sqrt{n}\sigma} \leq \frac{S_n}{\sqrt{n}\sigma} \leq \frac{x}{\sqrt{n}\sigma}\right\} \\
&= 1 - \left(\mathbb{P}\left\{\frac{S_n}{\sqrt{n}\sigma} \leq \frac{x}{\sqrt{n}\sigma}\right\} - \mathbb{P}\left\{\frac{S_n}{\sqrt{n}\sigma} < -\frac{x}{\sqrt{n}\sigma}\right\}\right) \\
&= 1 - \left(\mathbb{P}\left\{\frac{S_n}{\sqrt{n}\sigma} \leq \frac{x}{\sqrt{n}\sigma}\right\} - \left(1 - \mathbb{P}\left\{\frac{S_n}{\sqrt{n}\sigma} \geq -\frac{x}{\sqrt{n}\sigma}\right\}\right)\right) \\
&\approx 1 - \left(\mathbb{P}\left\{\frac{S_n}{\sqrt{n}\sigma} \leq \frac{x}{\sqrt{n}\sigma}\right\} - \left(1 - \mathbb{P}\left\{\frac{S_n}{\sqrt{n}\sigma} \leq \frac{x}{\sqrt{n}\sigma}\right\}\right)\right) \\
&= 2\left(1 - \mathbb{P}\left\{\frac{S_n}{\sqrt{n}\sigma} \leq \frac{x}{\sqrt{n}\sigma}\right\}.\right)
\end{aligned}$$

Since we expect $\frac{S_n}{\sqrt{n}\sigma}$ to be normally distributed, letting F denote the distribution function of $\mathcal{N}(0, 1)$,

$$\mathbb{P}\left\{|S_n| > x\right\} \approx 2\left(1 - F\left(\frac{x}{\sqrt{n}\sigma}\right)\right).$$

For example, if $n = 10$ and $x = 1$, then

$$\mathbb{P}\left\{|S_{10}| > 1\right\} \approx 2\left(1 - F\left(\frac{\sqrt{12}}{\sqrt{10}}\right)\right) \approx 0.27.$$

3 Confidence intervals

The discussion in this section is very hand wavy, so take it with a grain of salt.

Consider conducting a sequence of trials represented as i.i.d. random variables $(X_j)_j$. For example, we may repeatedly toss an unfair coin repeatedly in order to determine its distribution. In practice, we will not have access to the variance of X_1.

Letting $S_n = X_1 + \cdots + X_n$, and consider the random variable

$$Z_n = \frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mu\right).$$

Then,

$$\mathbb{P}(-z \leq Z_n \leq z) = \mathbb{P}\left(\mu - z\frac{\sigma}{\sqrt{n}} \leq \frac{S_n}{n} \leq \mu + z\frac{\sigma}{\sqrt{n}}\right).$$

Now, fix $\alpha \in [0,1]$. Pick z such that

$$\mathbb{P}(-z \leq Z_n \leq z) = 1 - \alpha.$$

Since $Z_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$,

$$\begin{aligned}
\mathbb{P}(-z \leq Z_n \leq z) &= \mathbb{P}(Z_n \leq z) - \mathbb{P}(Z_n < -z) \\
&= \mathbb{P}(Z_n \leq z) - (1 - \mathbb{P}(Z_n \geq -z)) \\
&\approx \mathbb{P}(Z_n \leq z) - (1 - \mathbb{P}(Z_n \leq z)) \\
&= 2\mathbb{P}(Z_n \leq z) - 1
\end{aligned}$$

where we have used the approximation

$$\mathbb{P}(Z_n \geq -z) \approx \mathbb{P}(Z_n \leq z)$$

which is a consequence of Z_n being nearly normal. Therefore,

$$\mathbb{P}(Z_n \leq z) \approx 1 - \frac{\alpha}{2}$$

and hence

$$z \approx F^{-1}\left(1 - \frac{\alpha}{2}\right).$$

This tells us that our sample mean S_n/n is "very likely" to be within

$$\pm F^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}$$

of the actual mean. However, this estimate is not so useful since we may not have access to σ. So instead, we make another approximation, substituting the sample variance:

$$F^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{\sigma_n}{\sqrt{n}}$$

where

$$\sigma_n = \frac{1}{n}\sum_{j=1}^{n}(X_j - S_n)^2.$$

Example 3.1. We toss a coin 100 times. The result of the j-th toss is X_j. Suppose we observe 53 heads so that

$$\frac{S_n(\omega)}{n} = 0.53 \qquad \text{and} \qquad \sigma_n^2 \approx 0.25$$

where we have written $S_n(\omega)$ above to stress that 0.53 is an observation. Let $\alpha = 0.05$, corresponding to a confidence interval of $1 - \alpha = 0.95$. Then,

$$F^{-1}\left(1 - \frac{\alpha}{2}\right) = F^{-1}(0.975) = 1.96.$$

Moreover,

$$1.96 \cdot \frac{\sigma_n}{\sqrt{n}} \approx 0.1$$

The confidence interval is

$$[0.53 - 0.1, 0.53 + 0.1] = [0.43, 0.63].$$

4 Lyapunov's CLT

The problem with the classical CLT is that it requires the variables to be identically distributed. Lyapunov's CLT relaxes this assumption:

Proposition 4.1. *Let $(X_j)_j$ be a sequence independent, mean zero random variables such that X_j^3 is integrable. Let $\sigma_j^2 = \mathrm{Var}(X_j)$, $\hat{\gamma}_j = \mathbb{E}[X_j^3]$, and $\gamma_j = \mathbb{E}[|X_j|^3]$. Let $S_n = X_1 + \cdots + X_n$ and $s_n = \sigma_1^2 + \cdots + \sigma_n^2$. Suppose that*

$$\frac{\gamma_1 + \cdots + \gamma_n}{s_n^3} \to 0 \qquad \text{as} \qquad n \to \infty.$$

Then,

$$\frac{S_n}{s_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \qquad \text{as} \qquad n \to \infty.$$