# PT: Lecture #19

November 12

**Recall:** The LLN (in its weak or strong form) is fundamentally about the mean of an average. That is, we consider e.g. $\{X_i\}_{i=1}^n$ iid random variables with finite expected value and variance $\mu, \sigma^2 \geq 0$. Then the LLN essentially says that

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\text{average of the random variables}} \longrightarrow \underbrace{\mu}_{\text{expectation of each } X_i} \quad \text{as } n \to \infty.$$

- This is central to statistical inference because it says if I observe lots of random samples from $\{X_i\}_{i=1}^n$ and take the average of my observations $\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$, this should be close to $\mu$ $(\mathbb{E}(X_i))$ as $n \to \infty$.

- In practice, this is often insufficient. We need to know more than

"$\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\text{random}}$ concentrates around $\underbrace{\mu}_{\text{deterministic}}$"

we need to know

"What does the probability distribution of $\frac{1}{n}\sum_{i=1}^{n} X_i$ look like"

- This is the concern of the central limit Theorem (CLT).

- We will start with a case that we understand well: binomial r.v.'s, a.k.a. sums of independent Bernoulli trials.

- Let $X_i \sim \text{Bern}(p)$, so that $\underbrace{\sum_{i=1}^{n} X_i}_{\| \; \therefore \; S_n} \sim \text{Bin}(n,p)$.

We know from LLN that $\frac{1}{n} S_n \to \mu$ as $n \to \infty$. So, we would expect at an intuitive level that $S_n \to n\mu$ as $n \to \infty$, but this is not true precisely, because $S_n - n\mu$ is getting large as $n \to \infty$ in absolute terms.

- What we can say is that in relative terms $S_n - n\mu$ is getting small. The crucial question for CLT is: how to compute the relative error of $S_n - n\mu$?

Defn: For $S_n \sim \text{Bin}(n,p)$, let $S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}$ be the normalized sum.

Proposition: For all $n$, (i) $\mathbb{E}(S_n^*) = 0$

(ii) $\text{Var}(S_n^*) = 1$

Prof: (i) Notice that $\mathbb{E}(S_n^*) \overset{\text{defn}}{=} \mathbb{E}\left(\frac{S_n - np}{\sqrt{np(1-p)}}\right)$

$$= \frac{1}{\sqrt{np(1-p)}} \left[ \mathbb{E}(S_n) - np \right]$$

$$= \frac{1}{\sqrt{np(1-p)}} \left[ np - np \right]$$

$$= \frac{1}{\sqrt{np(1-p)}} \left[ 0 \right]$$

$$= 0. \checkmark$$

(ii) Similarly, $Var(S_n^*) = Var\left( \frac{S_n - np}{\sqrt{np(1-p)}} \right)$

$$= \frac{1}{np(1-p)} Var(S_n - np)$$

$$= \frac{1}{np(1-p)} Var(S_n)$$

$$= \frac{1}{np(1-p)} \cdot np(1-p)$$

$$= 1. \checkmark \quad \blacksquare$$

Note: Translating $S_n$ by the (deterministic) constant $np$ does not change the variance.

- So, $S_n^*$ is mean 0, variance 1. The content of our first CLT is that as $n \to \infty$, $S_n^*$ is distributed as a Gaussian with mean 0 and variance 1.
- In order to show this, we can show that $P(S_n = k) \approx \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left( \frac{-[k-np]^2}{2np(1-p)} \right)$, since this is essentially the non-standardized variation of the above.

**Theorem (De Moivre-Laplace CLT):** Let $S_n \sim Bin(n,p)$. Then for all $k$ s.t. $\frac{k}{n} \to p$,

$$\lim_{n \to \infty} \frac{P(S_n = k)}{\frac{1}{\sqrt{2\pi np(1-p)}} \cdot \exp\left( -\frac{[k-np]^2}{np(1-p)} \right)} = 1$$

**Proof:** We will use Stirling's approximation. Let $a_n \simeq b_n$ denote $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

$$\mathbb{P}(S_n = k) \overset{\text{def}_n}{=} \binom{n}{k} \rho^k (1-\rho)^{n-k}$$

$$= \frac{n!}{k!(n-k)!} \rho^k (1-\rho)^{n-k}$$

**Stirling's Approximation**

$$n! \simeq n^n e^{-n} \cdot \sqrt{2\pi n}$$

$$\overset{\text{Stirling}}{\simeq} \frac{n^n \cdot e^{-n} \cdot \sqrt{2\pi n}}{k^k \cdot e^{-k} \cdot \sqrt{2\pi k} \cdot (n-k)^{n-k} \cdot e^{-(n-k)} \cdot \sqrt{2\pi(n-k)}} \cdot \rho^k \cdot (1-\rho)^{n-k}$$

$$= \sqrt{\frac{n}{2\pi k(n-k)}} \cdot \frac{n^n}{k^k \cdot (n-k)^{n-k}} \cdot \rho^k \cdot (1-\rho)^{n-k}$$

$$= \sqrt{\frac{n}{2\pi k(n-k)}} \cdot \left(\frac{n\rho}{k}\right)^k \cdot \left(\frac{n[1-\rho]}{n-k}\right)^{n-k}$$

$$= \sqrt{\frac{1}{2\pi n \cdot \frac{k}{n} \cdot (1-\frac{k}{n})}} \cdot \left(\frac{n\rho}{k}\right)^k \cdot \left(\frac{n[1-\rho]}{n-k}\right)^{n-k}$$

$$\simeq \underbrace{\sqrt{\frac{1}{2\pi \cdot n \cdot \rho \cdot (1-\rho)}}}_{= C_{n,\rho}} \cdot \left(\frac{n\rho}{k}\right)^k \cdot \left(\frac{n[1-\rho]}{n-k}\right)^{n-k}$$

$$= C_{n,\rho} \cdot \exp\left[\ln\left(\frac{n\rho}{k}\right)^k + \ln\left(\frac{n[1-\rho]}{n-k}\right)^{n-k}\right]$$

$$= C_{n,\rho} \cdot \exp\left[-k \ln\left(\frac{k}{n\rho}\right) + (k-n)\ln\left(\frac{n-k}{n[1-\rho]}\right)\right]$$

$$= C_{n,p} \cdot \exp\left( -k \ln\left( \frac{np + x\sqrt{np(1-p)}}{np} \right) + (k-n) \ln\left( \frac{n - np - x\sqrt{np(1-p)}}{n[1-p]} \right) \right), \text{ noting } k = np + x\sqrt{np(1-p)}$$

$$= C_{n,p} \cdot \exp\left( -k \ln\left( 1 + x\sqrt{\frac{1-p}{np}} \right) + (k-n) \ln\left( 1 - x\sqrt{\frac{p}{n(1-p)}} \right) \right)$$

Taylor

$$= C_{n,p} \cdot \exp\left( -k \left( x\sqrt{\frac{1-p}{np}} - \frac{1}{2}x^2\left(\frac{1-p}{np}\right) + \dots \right) + (k-n)\left( -x\sqrt{\frac{p}{n(1-p)}} - x^2\left(\frac{p}{n(1-p)}\right) - \dots \right) \right)$$

expanding $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$

$$= C_{n,p} \cdot \exp\left( -\left[np + x\sqrt{np(1-p)}\right]\left( x\sqrt{\frac{1-p}{np}} - \frac{1}{2}x^2\left(\frac{1-p}{np}\right) + \dots \right) + \left(np + x\sqrt{np(1-p)} - n\right)\left( -x\sqrt{\frac{p}{n(1-p)}} - x^2\left(\frac{p}{n(1-p)}\right) - \dots \right) \right)$$

$$= C_{n,p} \cdot \exp\left( \left[ -np - x\sqrt{np(1-p)} \right]\left( x\sqrt{\frac{1-p}{np}} - \frac{1}{2}x^2\left(\frac{1-p}{np}\right) + \dots \right) - \left[ n[1-p] - x\sqrt{np(1-p)} \right]\left( -x\sqrt{\frac{p}{n(1-p)}} - x^2\left(\frac{p}{n(1-p)}\right) - \dots \right) \right)$$

$$= C_{n,p} \cdot \exp\left( \left[ -x\sqrt{np(1-p)} + \frac{1}{2}x^2(1-p) - x^2(1-p) + \text{H.O.T. in } n^{-1} \right] + \left[ x\sqrt{np(1-p)} + \frac{1}{2}x^2 p - x^2 p - \text{H.O.T. in } n^{-1} \right] \right)$$

$$= C_{n,p} \cdot \exp\left( -\frac{1}{2}x^2(1-p) - \frac{1}{2}x^2(p) + \text{H.O.T.} \right)$$

$k = np + x\sqrt{np(1-p)} \Rightarrow x = \frac{nk - np}{\sqrt{np(1-p)}}$

$$\approx C_{n,p} \cdot \exp\left( -\frac{1}{2}x^2 \right) \quad \text{As } n \to \infty, \text{ the H.O.T.} \to 0$$

$$= \frac{1}{\sqrt{2\pi np(1-p)}} \cdot \exp\left( \frac{-(np-k)^2}{2np(1-p)} \right), \text{ as claimed.} \blacksquare$$

Remark: While tedious, this result is elementary in the sense that it doesn't need complex analysis; the big tool is Stirling's approximation. Compared to more general CLT, this result is quite old, ~1750.

Corollary: With $S_n$ as above in De Moivre-Laplace CLT, for all $a < b$,

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-z^2/2\right) dz.$$