13: Limit Laws

PSTAT 120A: Summer 2022

Ethan P. Marzban

July 25, 2022

University of California, Santa Barbara

Where We've Been

- Axioms of Probability, Probability Spaces, Counting
- Conditional Probabilities, independence, etc.
- Basics of Random Variables (classification, p.m.f., c.m.f., moments)
- Discrete Distributions
- Continuous Distributions
- Transformations of Random Variables
- Double Integrals
- Random Vectors and the basics of multivariate probability
- Independence of random variables, and covariance/correlation
- Sums of Random Variables; Indicators
- Moment Generating Functions
- Tail Bounds

Law of Large Numbers

Convergence in Distribution

Definition: Convergence in Distribution

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sequence $\{X_n\}$ of random variables with corresponding c.d.f.'s $F_n(x) := F_{X_n}(x)$, and another random variable X with c.d.f $F_X(x)$. We say that the sequence $\{X_n\}$ **converges in distribution** to X if we have pointwise convergence of the c.d.f.'s. In other words:

$$F_n(x) \to F_X(x) \ \forall x \text{ for which } F_n(x) \text{ and } F(x) \text{ are continuous}$$

We denote convergence in distribution using the $\rightsquigarrow$ symbol: i.e. $X_n \rightsquigarrow X$

Degenerate Random Variables

Definition

A [discrete] random variable X is said to be **degenerate** if there exists a point x' in its state space S_X for which $\mathbb{P}(X = x') = 1$.

- Note, in this way, that constants can be viewed as random variables, albeit degenerate ones. As such, it makes perfect sense to say that a sequence of random variables $\{X_n\}$ converges in distribution to a constant.

Convergence in Probability

Definition: Convergence in Probability

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sequence $\{X_n\}$ of random variables, and another random variable X. We say that the sequence $\{X_n\}$ **converges in probability** to X if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$$

We denote convergence in distribution using the $\xrightarrow{\text{p}}$ symbol: i.e. $X_n \xrightarrow{\text{p}} X$

- Though we will not prove it, convergence in probability is a stronger condition than convergence in distribution. That is, if $X_n \xrightarrow{\text{p}} X$ then $X_n \rightsquigarrow X$ whereas the converse is not necessarily true.

Law of Large Numbers

Theorem: Law of Large Numbers

Given an i.i.d. sequence of random variables $\{X_i\}_{i=1}^{\infty}$ with common mean μ and variance σ^2, then

$$\overline{X}_n \xrightarrow{\text{p}} \mu$$

where $\overline{X}_n := n^{-1} \sum_{i=1}^{n} X_i$ denotes the sample mean. Phrased differently, the LLN states that, for any fixed $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left(|\overline{X}_n - \mu| \geq \varepsilon\right) = 0$$

Proof.

- First note that, by a previous homework problem, $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathrm{Var}(\overline{X}_n) = \sigma^2/n$.
- Thus, Chebyshev's inequality tells us

$$\mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

- Additionally, probabilities are definitionally nonnegative meaning we have

$$0 \leq \mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

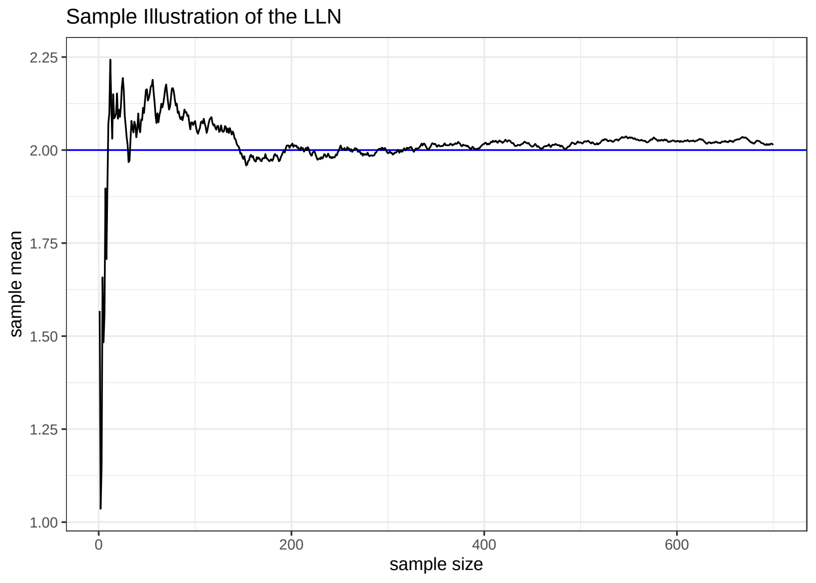
- Since $\sigma^2 < \infty$ is fixed, the upper bound above goes to 0 as $n \to \infty$. Since $0 \to 0$ as $n \to \infty$, we utilize the Squeeze Theorem to conclude that

$$\lim_{n \to \infty} \mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) = 0$$

which shows that $\overline{X}_n \xrightarrow{\mathrm{p}} \mu$.

□

Heuristic Illustration of the LLN



Central Limit Theorem

Central Limit Theorem

Theorem: Central Limit Theorem

Given a sequence of i.i.d. random variables $\{X_i\}_{i=1}^{\infty}$ with common mean μ and variance σ^2, define $S_n := \sum_{i=1}^{n} X_i = X_1 + \cdots + X_n$. Additionally, let Z be a random variable such that $Z \sim \mathcal{N}(0,1)$. Then

$$\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \right) \rightsquigarrow Z$$

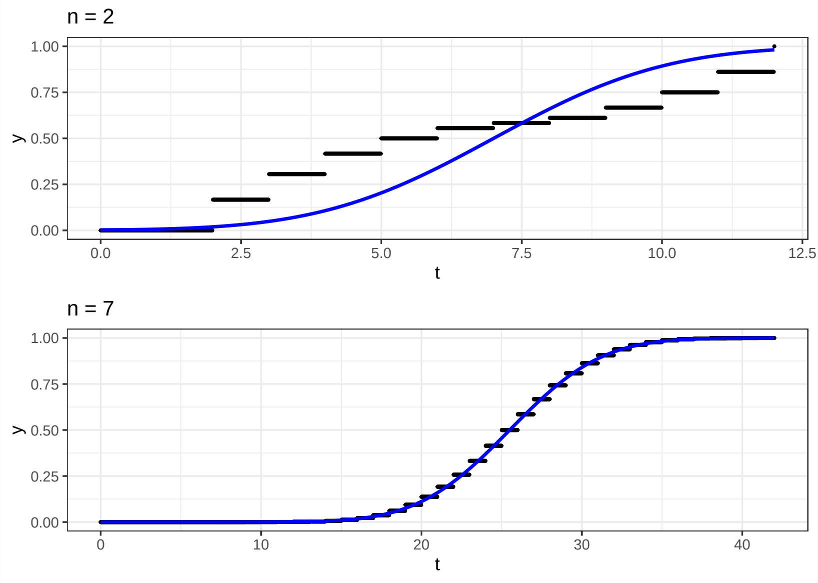
which is often abbreviated

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

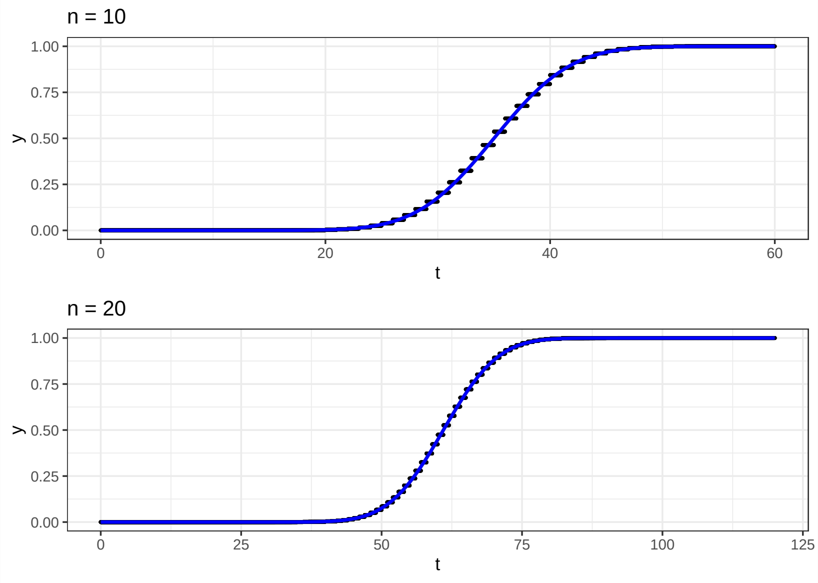
or, more accurately,

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}z^2}\, \mathrm{d}z$$

Heuristic Illustration of the CLT



Heuristic Illustration of the CLT



Heuristic Illustration of the CLT

Shiny App!

Proof Sketch

Proof.

- Our proof shall utilize MGF's. Define

$$Y_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Then,

$$\begin{aligned} M_{Y_n}(t) = \mathbb{E}[e^{tY_n}] &= \mathbb{E}\left[\exp\left\{t \cdot \frac{S_n - n\mu}{\sigma\sqrt{n}}\right\}\right] \\ &= \mathbb{E}\left[\exp\left\{\frac{t}{\sigma\sqrt{n}}\right\}\sum_{i=1}^{n}(X_i - \mu)\right] \\ &= \mathbb{E}\left[\prod_{i=1}^{n} e^{\frac{t}{\sigma\sqrt{n}}(X_i - \mu)}\right] \end{aligned}$$

where we have utilized the definition $S_n := \sum_{k=1}^{n} X_i$ to go from the first line to the second.

□

Proof Sketch

Proof.

- At this point, we utilize the independence of the X_i's to interchange the expectation and the product operator:

$$M_{Y_n}(t) = \mathbb{E}\left[\prod_{i=1}^{n} e^{\frac{t}{\sigma\sqrt{n}}(X_i-\mu)}\right]$$
$$= \prod_{i=1}^{n} \mathbb{E}\left[e^{\frac{t}{\sigma\sqrt{n}}(X_i-\mu)}\right]$$

□

Proof Sketch

Proof.

- Now, recall that we know the MacLaurin Series Expansion of e^y:

$$e^y = \sum_{k=0}^{\infty} \frac{y^k}{k!} = 1 + x + \frac{1}{2}x^2 + \cdots$$

We shall apply a second-order Taylor Series Expansion to the quantity inside our expectation above:

$$\begin{aligned}
M_{Y_n}(t) &= \prod_{i=1}^{n} \mathbb{E}\left[e^{\frac{t}{\sigma\sqrt{n}}(X_i-\mu)}\right] \\
&= \prod_{i=1}^{n} \mathbb{E}\left[1 + \frac{t}{\sigma\sqrt{n}}(X_i - \mu) + \frac{t^2}{2\sigma^2 n}(X_i - \mu)^2 + \cdots\right] \\
&= \prod_{i=1}^{n} \left[1 + \frac{t}{\sigma\sqrt{n}} \cancelto{0}{\mathbb{E}[X_i - \mu]} + \frac{t^2}{2\sigma^2 n} \cancelto{\sigma^2}{\mathbb{E}[(X_i - \mu)^2]} + \cdots\right] \\
&\approx \prod_{i=1}^{n} \left[1 + \frac{t^2}{2n}\right] = \left(1 + \frac{t^2}{2n}\right)^n
\end{aligned}$$

□

Proof Sketch

Proof.

- Our final task will be to take the limit as $n \to \infty$. First recall that

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$$

meaning

$$\begin{aligned}\lim_{n\to\infty} M_{Y_n}(t) &= \lim_{n\to\infty}\left[\left(1+\frac{t^2}{2n}\right)^n\right]\\ &= \lim_{n\to\infty}\left(1+\frac{\left(\frac{t^2}{2}\right)}{n}\right)^n = e^{\frac{t^2}{2}}\end{aligned}$$

which we recognize as the MGF of the $\mathcal{N}(0,1)$ distribution.

□

Comments on the Proof Sketch

- Now, the above is just a proof *sketch*. Firstly, we would need to examine the "$\approx$" signs more carefully. Additionally, we would need some sort of theorem that guarantees that the limit of an MGF will accurately give the limiting distribution. But, these are considerations for a future class; for the purposes of this class, I mainly wanted to illustrate how the basic idea of the proof relates to material we've seen throughout this class!

- By the way, the idea of examining the "long-term" behavior of distributions (i.e. when n is large) relates heavily to a field of statistics called **asymptotics**.

How Large is Large?

- Now, practically speaking, we are almost never afforded the luxury of an infinite sample size. So, a natural question that might arise is: under what *practical* conditions does the CLT give us an approximation that is reasonably close to the truth?
- Admittedly, there isn't a single agreed-upon set of cutoffs/criteria! I shall adopt a relatively simple one: $n \geq 25$. (In other words, if $n < 25$, then the CLT may not be a good idea)

Example

Every night, Melinda adds some money to her piggy bank. The amount she adds on any given day averages \$5 with a standard deviation of \$2. What is the probability that after a month (30 days) the total amount of money in Melinda's piggy bank will exceed \$155?

Sample Size Example

Suppose we have a sequence of i.i.d. random variables $\{X_i\}_{i=1}^{\infty}$ with common mean μ and variance σ^2. Find the sample size n that ensures the sample mean $\overline{X}_n$ exceeds some fixed value c with probability p.

A New Theorem… Or Is It?

Leadup

- Let $X \sim \text{Bin}(n, p)$. Recall how we saw that X can be expressed as the sum of n independent indicator random variables $\mathbb{1}_j$, where

$$\mathbb{1}_j = \begin{cases} 1 & \text{if } j^{\text{th}} \text{ trial resulted in success} \\ 0 & \text{otherwise} \end{cases}$$

- Hm, sums of many random variables... sounds like a potential application of the CLT!
- That's right; we can actually use the CLT to approximate a Binomial Distribution with a Normal distribution!

De Moivre-Laplace Theorem

Theorem: De Moivre-Laplace Theorem

Let $0 < p < 1$ be a fixed number, and suppose $S_n \sim \mathrm{Bin}(n, p)$. Then

$$\left(\frac{S_n - np}{\sqrt{np(1-p)}} \right) \rightsquigarrow \mathcal{N}(0, 1)$$

- In many ways, this is just an application of the CLT to the Binomial Theorem (after noting that a Binomially distributed random variable can be written as the sum of n i.i.d. random variables).
- We often call this the "Normal Approximation to the Binomial."
- Now, just as with the CLT, we need to be a bit careful about how large n should be in order for the approximation to be good. As a general rule of thumb, the De Moivre-Laplace Theorem works for a $\mathrm{Bin}(n, p)$ distribution where:
  (1) n is large (say, $n \geq 25$)
  (2) $np \geq 5$
  (3) $np(1-p) \geq 5$

  (though, again, there admittedly isn't a single agreed-upon cutoff for any of these values.)

Another Approximation

- What do we do if n is large, but, say, np is small?
- Well, perhaps you recall that we encountered limits in a Binomial setting once long ago... That's right; in the context of the Poisson Process!
- Specifically, we showed that if $X_n \sim \text{Bin}(n, p)$ we have, for sufficiently large n,

$$\mathbb{P}(X_n = k) \approx e^{-(np)} \cdot \frac{(np)^k}{k!}$$

- This gives us yet another approximation to the Binomial distribution; this time, when p is small!

Theorem

If $X \sim \text{Bin}(n, p)$, then the distribution of X is well-approximated by the $\text{Pois}(np)$ distribution provided that:

(1) n is large ($n \geq 25$)

(2) p is small ($p \leq 0.05$)

(3) np is small ($np < 5$)