

Lecture 10: Laws of Large Numbers and the Central Limit Theorem

EECS 126 (UC Berkeley)

Spring 2026 — February 19, 2026

Reading: Bertsekas & Tsitsiklis, Ch. 5; supplementary notes on convergence.

These are my personal notes from Professor Kannan Ramchandran's EECS 126 (Probability and Random Processes) course at UC Berkeley.

Topics: the weak law of large numbers (WLLN), convergence in probability, almost sure convergence and the strong law of large numbers (SLLN), and the central limit theorem (CLT).

10.1 Recap: Concentration Inequalities

The following bounds were derived in Lecture 9; we restate the ones used below.

Theorem 10.1 (Markov's inequality)

If $X \geq 0$ and $a > 0$, then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \quad (10.1)$$

Theorem 10.2 (Chebyshev's inequality)

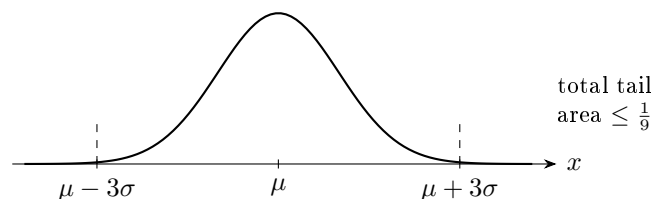
If X has mean μ and variance σ^2 , then for every $c > 0$,

$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\text{Var}(X)}{c^2} = \frac{\sigma^2}{c^2}. \quad (10.2)$$

With $c = k\sigma$:

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}. \quad (10.3)$$

The left side is the probability that X deviates from its mean by at least c . For $k = 3$, the two tails beyond $\mu \pm 3\sigma$ have total probability at most $\frac{1}{9}$, for *any* distribution with finite variance.



Theorem 10.3 (Chernoff bound)

For every b ,

$$\mathbb{P}(X \geq b) \leq \min_{s>0} e^{-sb} M_X(s), \quad M_X(s) = \mathbb{E}[e^{sX}]. \quad (10.4)$$

Example 10.1 (Chernoff bound for a Gaussian tail)

Let $X \sim \mathcal{N}(0, \sigma^2)$, so $M_X(s) = e^{\sigma^2 s^2/2}$ (Lecture 8). With $b = k\sigma$,

$$\mathbb{P}(X \geq k\sigma) \leq \min_{s>0} e^{-sk\sigma} e^{\sigma^2 s^2/2} = \min_{s>0} e^{h(s)}, \quad h(s) = \frac{\sigma^2 s^2}{2} - k\sigma s. \quad (10.5)$$

Minimize the exponent:

$$h'(s) = \sigma^2 s - k\sigma = 0 \implies s^* = \frac{k}{\sigma}, \quad h(s^*) = \frac{\sigma^2}{2} \cdot \frac{k^2}{\sigma^2} - k\sigma \cdot \frac{k}{\sigma} = \frac{k^2}{2} - k^2 = -\frac{k^2}{2}. \quad (10.6)$$

Therefore

$$\mathbb{P}(X \geq k\sigma) \leq e^{-k^2/2}. \quad (10.7)$$

Comparison at $k = 3$ (two-sided, $\mathbb{P}(|X| \geq 3\sigma)$):

Chebyshev	$\frac{1}{k^2} = \frac{1}{9}$	≈ 0.111
Chernoff	$2e^{-k^2/2} = 2e^{-4.5}$	≈ 0.0222
Exact	$2(1 - \Phi(3))$	≈ 0.0027

Chebyshev uses only the variance, so it holds for every distribution and is loose for the Gaussian. Chernoff uses the whole MGF and is much tighter.

10.2 The Weak Law of Large Numbers

Let $X_1, X_2, \dots, X_n$ be i.i.d. with mean μ and variance $\sigma^2 < \infty$. The **sample mean** (empirical mean estimate) is

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i. \quad (10.8)$$

Theorem 10.4 (Weak law of large numbers)

For every $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) \longrightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (10.9)$$

The probability that the sample mean deviates from the population mean by at least ε vanishes as $n \rightarrow \infty$.

Proof. Compute the mean and variance of M_n , then plug into Chebyshev. By linearity,

$$\mathbb{E}[M_n] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} \cdot n\mu = \mu. \quad (10.10)$$

By $\text{Var}(aX) = a^2 \text{Var}(X)$ and independence (Lecture 4),

$$\text{Var}(M_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \left(\frac{1}{n}\right)^2 \text{Var}\left(\sum_{i=1}^n X_i\right) = \left(\frac{1}{n}\right)^2 \cdot n\sigma^2 = \frac{\sigma^2}{n}. \quad (10.11)$$

By Chebyshev (10.2) with $X = M_n$ and $c = \varepsilon$,

$$\mathbb{P}(|M_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(M_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (10.12)$$

□

Formally, for all $\varepsilon, \delta > 0$ there exists $n(\varepsilon, \delta)$ such that

$$\mathbb{P}(|M_n - \mu| \geq \varepsilon) \leq \delta \quad \text{for all } n \geq n(\varepsilon, \delta). \quad (10.13)$$

From (10.12), $\frac{\sigma^2}{n\varepsilon^2} \leq \delta \iff n \geq \frac{\sigma^2}{\varepsilon^2\delta}$, so $n(\varepsilon, \delta) = \frac{\sigma^2}{\varepsilon^2\delta}$ samples suffice. For example, with $\sigma^2 = 1$, $\varepsilon = 0.1$, $\delta = 0.05$: $n \geq \frac{1}{(0.01)(0.05)} = 2000$.

What are ε and δ capturing? I had a question of what ε captures in this statement.

- ε is the **accuracy**: how close M_n must be to μ .
- δ is the **confidence**: the probability of a deviation of at least ε is at most δ . (It bounds a probability, not the deviation itself.)
- $n(\varepsilon, \delta)$ is the number of samples needed for a given accuracy and confidence.

In words: for $n \geq n(\varepsilon, \delta)$, $|M_n - \mu| < \varepsilon$ with probability at least $1 - \delta$.

10.3 Convergence in Probability

Definition 10.1 (Convergence in probability)

A sequence $Y_1, Y_2, \dots$ **converges in probability** to a constant a , written $Y_n \xrightarrow{p} a$, if for every $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - a| \geq \varepsilon) = 0. \quad (10.14)$$

The WLLN says exactly that $M_n \xrightarrow{p} \mu$.

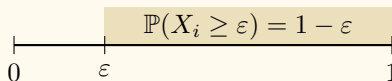
Example 10.2 (Minimum of uniforms)

Let $X_1, X_2, \dots$ be i.i.d. Uniform $[0, 1]$ and $Y_n = \min(X_1, \dots, X_n)$. Does Y_n converge in probability, and to what?

As we draw more samples, the smallest one should approach 0. Fix $0 < \varepsilon < 1$. Since $Y_n \geq 0$,

$$\begin{aligned} \mathbb{P}(|Y_n - 0| \geq \varepsilon) &= \mathbb{P}(Y_n \geq \varepsilon) = \mathbb{P}(\min(X_1, \dots, X_n) \geq \varepsilon) \\ &= \mathbb{P}(X_1 \geq \varepsilon, X_2 \geq \varepsilon, \dots, X_n \geq \varepsilon) \quad (\text{all must be } \geq \varepsilon) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \geq \varepsilon) \quad (\text{independence}) \\ &= (1 - \varepsilon)^n \rightarrow 0 \quad \text{as } n \rightarrow \infty, \end{aligned} \quad (10.15)$$

since $0 < 1 - \varepsilon < 1$. (For $\varepsilon \geq 1$ the probability is 0.) Therefore $Y_n \xrightarrow{p} 0$.



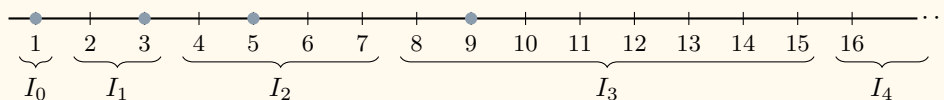
Example 10.3 (Discrete-time arrival process)

Time runs in slots $n = 1, 2, 3, \dots$, grouped into intervals

$$I_k = \{2^k, 2^k + 1, \dots, 2^{k+1} - 1\}, \quad k = 0, 1, 2, \dots, \quad (10.16)$$

so $I_0 = \{1\}$, $I_1 = \{2, 3\}$, $I_2 = \{4, \dots, 7\}$, $I_3 = \{8, \dots, 15\}$, and $|I_k| = 2^k$. There is exactly 1 arrival in each interval I_k , at a time chosen uniformly at random within I_k . Let

$$Y_n = \begin{cases} 1 & \text{if there is an arrival at time } n, \\ 0 & \text{otherwise.} \end{cases} \quad (10.17)$$



Each time in I_k is equally likely to hold the arrival, so

$$\begin{aligned} \mathbb{P}(Y_1 = 1) &= 1, \\ \mathbb{P}(Y_2 = 1) &= \mathbb{P}(Y_3 = 1) = \frac{1}{2}, \\ \mathbb{P}(Y_4 = 1) &= \dots = \mathbb{P}(Y_7 = 1) = \frac{1}{4}, \end{aligned} \quad (10.18)$$

and in general

$$\mathbb{P}(Y_n = 1) = \frac{1}{2^k} \quad \text{if } n \in I_k. \quad (10.19)$$

As $n \rightarrow \infty$, the interval index k containing n also goes to ∞ , so

$$\lim_{n \rightarrow \infty} \mathbb{P}(Y_n = 1) = \lim_{k \rightarrow \infty} \frac{1}{2^k} = 0. \quad (10.20)$$

For $0 < \varepsilon \leq 1$, $\mathbb{P}(|Y_n - 0| \geq \varepsilon) = \mathbb{P}(Y_n = 1) \rightarrow 0$ (and it is 0 for $\varepsilon > 1$), so $Y_n \xrightarrow{p} 0$.

Converging, yet arrivals keep coming. Initially I was wondering how Y_n can converge to 0 when, no matter which n I pick, infinitely many arrivals still come after it. This is why convergence in probability is a *weak* notion: it only says that at each fixed large n , an arrival at that particular time is unlikely. It says nothing about a single realization of the whole sequence, which still has $Y_n = 1$ once in every interval. This weakness belongs to convergence in probability, not to the WLLN itself.

10.4 Almost Sure Convergence and the Strong Law

Each outcome ω of the experiment produces a whole *realization*: an ordinary sequence of numbers $Y_1(\omega), Y_2(\omega), \dots$

Definition 10.2 (Almost sure convergence)

Y_n converges almost surely (a.s.) to Y , written $Y_n \xrightarrow{\text{a.s.}} Y$, if

$$\mathbb{P}\left(\left\{\omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega)\right\}\right) = 1. \quad (10.21)$$

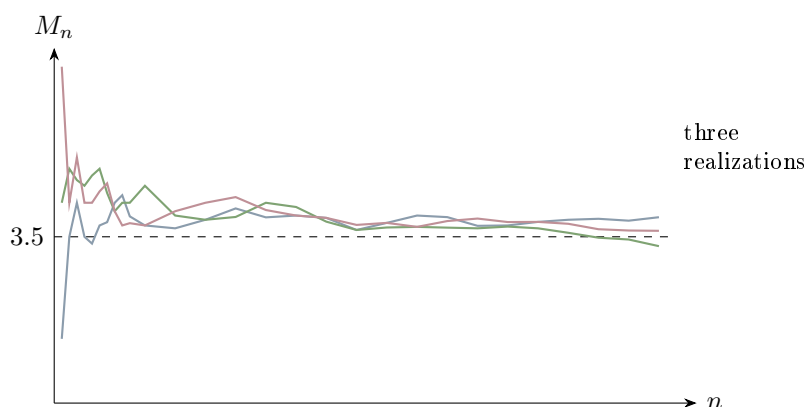
That is, the realization converges as an ordinary sequence for every outcome outside an event of probability 0.

Theorem 10.5 (Strong law of large numbers)

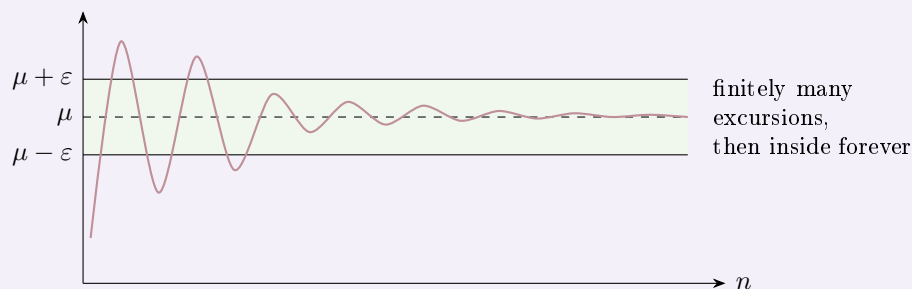
Let $\{X_n\}_{n \geq 1}$ be i.i.d. with mean μ . Then

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mu\right) = 1, \quad \text{i.e.} \quad M_n \xrightarrow{\text{a.s.}} \mu. \quad (10.22)$$

(The proof is beyond the scope of this course.) For fair die rolls, $\mu = \frac{1+2+\dots+6}{6} = \frac{21}{6} = 3.5$, and every realization of M_n (except on an event of probability 0) converges to 3.5:



Excursions: SLLN versus WLLN. Initially I was wondering whether the SLLN allows any excursions at all outside a band $\mu \pm \varepsilon$. It does: almost every realization leaves the band only *finitely many* times, and after some point stays inside forever. Convergence in probability (the WLLN) only requires that, at each fixed large n , few realizations are outside the band; a single realization may still leave it infinitely often.



Theorem 10.6 (Almost sure convergence implies convergence in probability)

If $Y_n \xrightarrow{\text{a.s.}} Y$, then $Y_n \xrightarrow{p} Y$. The converse is not necessarily true.

Proof. Fix $\varepsilon > 0$ and let

$$A_n = \bigcup_{m \geq n} \{|Y_m - Y| \geq \varepsilon\}, \quad (10.23)$$

the event that the realization leaves the ε -band at some time $m \geq n$. Then $A_1 \supseteq A_2 \supseteq \dots$. An outcome in every A_n leaves the band infinitely often, so its realization does not converge to Y ; hence $\mathbb{P}(\bigcap_n A_n) \leq \mathbb{P}(Y_n \not\rightarrow Y) = 0$. By continuity of probability for decreasing events (a consequence of countable additivity, Lecture 1), $\mathbb{P}(A_n) \rightarrow \mathbb{P}(\bigcap_n A_n) = 0$. Since $\{|Y_n - Y| \geq \varepsilon\} \subseteq A_n$,

$$\mathbb{P}(|Y_n - Y| \geq \varepsilon) \leq \mathbb{P}(A_n) \rightarrow 0. \quad (10.24)$$

□

The converse fails: the arrival process. Every realization has exactly one arrival in each interval I_k , so $Y_n(\omega) = 1$ for infinitely many n and $Y_n(\omega) = 0$ for infinitely many n . The sequence $Y_n(\omega)$ never converges, for *every* outcome ω , so $\mathbb{P}(Y_n \rightarrow 0) = 0$. Thus $Y_n \xrightarrow{p} 0$ but Y_n does not converge almost surely.

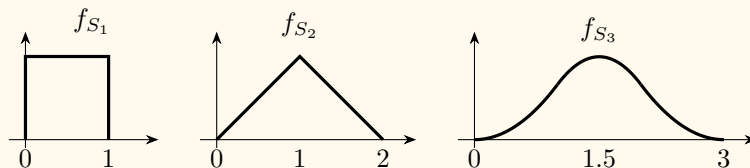
10.5 The Central Limit Theorem

Let $X_1, X_2, \dots$ be i.i.d. and $S_n = \sum_{i=1}^n X_i$. What happens to the distribution of S_n as n grows?

Example 10.4 (Sums of uniforms)

$X_i \sim \text{Uniform}[0, 1]$. f_{S_1} is a box and f_{S_2} is the triangle from Lecture 8. Convolving once more, $f_{S_3}(z) = \int_{z-1}^z f_{S_2}(x) dx$:

$$\begin{aligned} 0 \leq z \leq 1: & \int_0^z x dx = \frac{z^2}{2}, \\ 1 \leq z \leq 2: & \int_{z-1}^1 x dx + \int_1^z (2-x) dx = \frac{1-(z-1)^2}{2} + \left(2z - \frac{z^2}{2} - \frac{3}{2}\right) = \frac{-2z^2 + 6z - 3}{2}, \\ 2 \leq z \leq 3: & \int_{z-1}^2 (2-x) dx = \frac{(3-z)^2}{2}. \end{aligned} \quad (10.25)$$



f_{S_3} is symmetric about 1.5 and already looks bell-shaped.

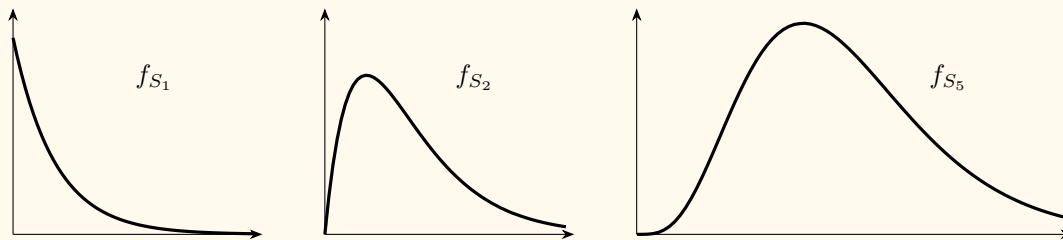
Example 10.5 (Sums of exponentials)

$X_i \sim \text{Exponential}(\lambda)$, so $f_{S_1}(x) = f_{X_1}(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. By convolution,

$$f_{S_2}(x) = \int_0^x \lambda e^{-\lambda u} \lambda e^{-\lambda(x-u)} du = \lambda^2 e^{-\lambda x} \int_0^x du = \lambda^2 x e^{-\lambda x}, \quad x \geq 0. \quad (10.26)$$

By induction, $f_{S_n}(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$:

$$f_{S_{n+1}}(x) = \int_0^x \frac{\lambda^n u^{n-1} e^{-\lambda u}}{(n-1)!} \lambda e^{-\lambda(x-u)} du = \frac{\lambda^{n+1} e^{-\lambda x}}{(n-1)!} \int_0^x u^{n-1} du = \frac{\lambda^{n+1} x^n e^{-\lambda x}}{n!}. \quad (10.27)$$



Even starting from a very skewed density, f_{S_n} becomes bell-shaped.

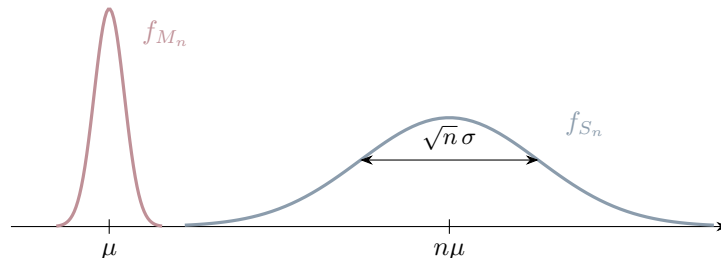
What happens to the mean and variance? By linearity and independence,

$$\mathbb{E}[S_n] = n\mu, \quad \text{Var}(S_n) = n\sigma^2. \quad (10.28)$$

The variance always grows, and the mean grows too unless $\mu = 0$, so f_{S_n} drifts and spreads out. What about the sample mean $M_n = \frac{S_n}{n}$?

$$\mathbb{E}[M_n] = \mu, \quad \text{Var}(M_n) = \frac{\sigma^2}{n} \longrightarrow 0, \quad (10.29)$$

so f_{M_n} collapses to a spike at μ . Neither shows the shape of the limit:



So M_n and S_n are both the wrong objects. The right normalization subtracts the mean and divides by the standard deviation of S_n :

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}. \quad (10.30)$$

Then

$$\mathbb{E}[Z_n] = \frac{\mathbb{E}[S_n] - n\mu}{\sqrt{n}\sigma} = \frac{n\mu - n\mu}{\sqrt{n}\sigma} = 0, \quad \text{Var}(Z_n) = \frac{\text{Var}(S_n)}{(\sqrt{n}\sigma)^2} = \frac{n\sigma^2}{n\sigma^2} = 1. \quad (10.31)$$

Standardizing does not make Z_n Gaussian. I had a question of whether the CLT just says that Z_n becomes Gaussian once it is normalized. $\mathbb{E}[Z_n] = 0$ and $\text{Var}(Z_n) = 1$ for every n , but for finite n , Z_n is generally *not* Gaussian (for uniform X_i , Z_n even has bounded support). The CLT is a statement about the *limit*: the CDF of Z_n approaches Φ as $n \rightarrow \infty$.

Definition 10.3 (Convergence in distribution)

Z_n **converges in distribution** to Z , written $Z_n \xrightarrow{d} Z$, if

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq x) = \mathbb{P}(Z \leq x) \quad (10.32)$$

at every x where the CDF of Z is continuous.

Theorem 10.7 (Central limit theorem)

Let $X_1, X_2, \dots$ be i.i.d. with mean μ and variance $0 < \sigma^2 < \infty$, and $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$. Then for every $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq x) = \Phi(x), \quad \Phi(x) = \mathbb{P}(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du, \quad (10.33)$$

where $Z \sim \mathcal{N}(0, 1)$ and u is a dummy variable. That is, $Z_n \xrightarrow{d} \mathcal{N}(0, 1)$.

Implications.

1. In the limit, the distribution of Z_n “wipes out” all information about the distribution of the X_i except the mean and the variance.
2. If there are a large number of independent small disturbances, their aggregate effect is approximately normally distributed. This is why the Gaussian is a good model for noise.

The finite-variance assumption is needed: without it, Z_n need not approach a Gaussian.

Proof via MGFs. Assume $M_{X_1}(s) < \infty$ for s in an interval around 0. Standardize each term: $Y_i = \frac{X_i - \mu}{\sigma}$, i.i.d. with $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}[Y_i^2] = \text{Var}(Y_i) = 1$. Then

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n}\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i. \quad (10.34)$$

By the MGF properties of Lecture 8 (scaling, then independent sums),

$$M_{Z_n}(s) = \mathbb{E}\left[e^{\frac{s}{\sqrt{n}} \sum_{i=1}^n Y_i}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{\frac{s}{\sqrt{n}} Y_i}\right] = \left(M_Y\left(\frac{s}{\sqrt{n}}\right)\right)^n. \quad (10.35)$$

Taylor-expand M_Y around 0, using $M_Y^{(k)}(0) = \mathbb{E}[Y^k]$ (Lecture 8):

$$M_Y(t) = M_Y(0) + M_Y'(0)t + \frac{M_Y''(0)}{2}t^2 + o(t^2) = 1 + \mathbb{E}[Y]t + \frac{\mathbb{E}[Y^2]}{2}t^2 + o(t^2) = 1 + \frac{t^2}{2} + o(t^2). \quad (10.36)$$

With $t = \frac{s}{\sqrt{n}}$,

$$M_Y\left(\frac{s}{\sqrt{n}}\right) = 1 + \frac{s^2}{2n} + o\left(\frac{1}{n}\right). \quad (10.37)$$

Take logarithms and use $\ln(1+x) = x + O(x^2)$ as $x \rightarrow 0$:

$$\ln M_{Z_n}(s) = n \ln \left(1 + \frac{s^2}{2n} + o\left(\frac{1}{n}\right) \right) = n \left(\frac{s^2}{2n} + o\left(\frac{1}{n}\right) \right) = \frac{s^2}{2} + o(1) \rightarrow \frac{s^2}{2}. \quad (10.38)$$

So $M_{Z_n}(s) \rightarrow e^{s^2/2}$, which is the MGF of $\mathcal{N}(0, 1)$ (Lecture 8). Finally, we use without proof the *continuity theorem*: if $M_{Z_n}(s) \rightarrow M_Z(s)$ for all s in an interval around 0, then $Z_n \xrightarrow{d} Z$. Hence $Z_n \xrightarrow{d} \mathcal{N}(0, 1)$. $\square$

This proof assumes the MGF exists near 0. The CLT holds with only a finite variance; that more general proof uses characteristic functions and is beyond the scope of this course.

End of Lecture 10.