Markov and Chebyshev Inequalities

Recall that a random variable \( X \) is called nonnegative if \( P(X \geq 0) = 1 \).

**Theorem 1 (Markov’s inequality)**

Let \( X \) be a nonnegative random variable with mean \( E(X) \). Then for any \( t > 0 \)

\[
P(X \geq t) \leq \frac{E(X)}{t}
\]

**Proof:** Assume \( X \) is continuous with pdf \( f \). Then \( f(x) = 0 \) if \( x < 0 \), so

\[
E(X) = \int_{0}^{\infty} xf(x) \, dx \geq \int_{t}^{\infty} xf(x) \, dx \geq t \int_{t}^{\infty} f(x) \, dx
\]

\[
= tP(X \geq t)
\]

If \( X \) is discrete, replace the integrals with sums. \( \square \)

**Example:** Suppose \( X \) is nonnegative and \( P(X > 10) = 1/5 \); show that \( E(X) \geq 2 \). Also, Markov’s inequality for \( |X| \).

The following result often gives a much sharper bound if the MGF of \( X \) is finite in some interval around zero.

**Theorem 3 (Chernoff’s bound)**

Let \( X \) be a random variable with MGF \( M_X(t) \). Then for any \( a \in \mathbb{R} \)

\[
P(X \geq a) \leq \min_{t > 0} e^{-at} M_X(t)
\]

**Proof:** Fix \( t > 0 \). Then we have

\[
P(X \geq a) = P(tX \geq ta) = P(e^{tx} \geq e^{ta}) \leq \frac{E(e^{tx})}{e^{ta}} \quad \text{(Markov’s inequality)}
\]

\[
= e^{-ta} M_X(t)
\]

Since this holds for all \( t > 0 \) such that \( M_X(t) < \infty \), it must hold for the \( t \) minimizing the upper bound. \( \square \)

Theorem 2 (Chebyshev’s inequality)

Let \( X \) be a random variable with finite variance \( \text{Var}(X) \). Then for any \( t > 0 \)

\[
P(|X - E(X)| \geq t) \leq \frac{\text{Var}(X)}{t^2}
\]

**Proof:** Apply Markov’s inequality to the nonnegative random variable

\[
Y = |X - E(X)|^2
\]

\[
P(|X - E(X)| \geq t) = P(|X - E(X)|^2 \geq t^2) \leq \frac{E(|X - E(X)|^2)}{t^2}
\]

\[
= \frac{\text{Var}(X)}{t^2} \quad \square
\]

**Example:** Chebyshev with \( t = k\sqrt{\text{Var}(X)} \) . . .
**Example:** Suppose $X \sim N(0, 1)$. Apply Chernoff’s bound to upper bound $P(X \geq a)$ for $a > 0$ and compare with the bounds obtained from Chebyshev’s and Markov’s inequalities.

- Recall that a random variable $X$ is a function $X : \Omega \to \mathbb{R}$ that maps any point $\omega$ in the sample space $\Omega$ to a real number $X(\omega)$.
- A random variable $X$ must satisfy the following: any subset $A$ of $\Omega$ in the form
  \[ A = \{ \omega \in \Omega : X(\omega) \in B \} \]
  for any “reasonable” $B \subset \mathbb{R}$ is an event. For example $B$ can be any set obtained by a countable union and intersection of intervals.
- Recall that a sequence of real numbers $\{x_n\}_{n=1}^{\infty}$ is said to converge to a limit $x \in \mathbb{R}$ (notation: $x_n \to x$) if for any $\epsilon > 0$ there exists $N$ such that $|x_n - x| < \epsilon$ for all $n \geq N$.

### Convergence of Random Variables

A *probability space* is a triple $(\Omega, \mathcal{A}, P)$, where $\Omega$ is a *sample space*, $\mathcal{A}$ is a collection of subsets of $\Omega$ called *events*, and $P$ is a probability measure on $\mathcal{A}$. In particular, the set of events $\mathcal{A}$ satisfies

1. $\Omega$ is an event
2. If $A \subset \Omega$ is an event, then $A^c$ is also an event
3. If $A_1, A_2, A_3, \ldots$ are events, then so is $\bigcup_{n=1}^{\infty} A_n$.

$P$ is a function from the collection of events $\mathcal{A}$ to $[0, 1]$ which satisfies the *axioms of probability*:

1. $P(A) \geq 0$ for all events $A \in \mathcal{A}$.
2. $P(\Omega) = 1$.
3. If $A_1, A_2, A_3, \ldots$ are mutually exclusive events (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$), then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

### Definitions (Modes of convergence)

Let $\{X_n\} = X_1, X_2, X_3, \ldots$ be a sequence of random variables defined in a probability space $(\Omega, \mathcal{A}, P)$.

(i) We say that $\{X_n\}$ converges to a random variable $X$ *almost surely* (notation: $X_n \xrightarrow{a.s.} X$) if

\[ P(\{\omega : X_n(\omega) \to X(\omega)\}) = 1 \]

(ii) We say that $\{X_n\}$ converges to a random variable $X$ *in probability* (notation: $X_n \xrightarrow{P} X$) if for any $\epsilon > 0$ we have

\[ P(|X_n - X| > \epsilon) \to 0 \text{ as } n \to \infty \]

(iii) If $r > 0$ we say that $\{X_n\}$ converges to a random variable $X$ *in rth mean* (notation: $X_n \xrightarrow{r.m.} X$) if

\[ E(|X_n - X|^r) \to 0 \text{ as } n \to \infty \]

(iv) Let $F_n$ and $F$ denote the cdfs of $X_n$ and $X$, respectively. We say that $\{X_n\}$ converges to a random variable $X$ *in distribution* (notation: $X_n \xrightarrow{d} X$) if

\[ F_n(x) \to F(x) \text{ as } n \to \infty \]

for any $x$ such that $F$ is continuous at $x$. 
Remarks:

- A very important case of convergence in rth mean is when \( r = 2 \). In this case, \( E(|X_n - X|^2) \to 0 \) as \( n \to \infty \) and we say that \( \{X_n\} \) converges in mean square to \( X \).
- Almost sure convergence is often called convergence with probability 1.
- Almost sure convergence, convergence in rth mean, and convergence in probability all state that \( X_n \) is eventually close to \( X \) (in different senses) as \( n \) increases.
- In contrast, convergence in distribution is only a statement about the closeness of the distribution of \( X_n \) to that of \( X \) for large \( n \).

Example: Sequence \( \{X_n\} \) such that \( F_n(x) = F(x) \) for all \( x \in \mathbb{R} \) and \( n = 1, 2, \ldots \), but \( P(|X_n - X| > 1/2) = 1 \) for all \( n \ldots \)

- Theorem 4

The following implications hold:

\[
\begin{align*}
X_n & \xrightarrow{a.s.} X \\
X_n & \xrightarrow{p} X \\
X_n & \xrightarrow{d} X
\end{align*}
\]

Remark: We will show that in general, \( X_n \xrightarrow{a.s.} X \) does not imply that \( X_n \xrightarrow{r.m.} X \), and also that \( X_n \xrightarrow{r.m.} X \) does not imply \( X_n \xrightarrow{a.s.} X \).

- Theorem 5

Convergence in rth mean implies convergence in probability; i.e., if \( X_n \xrightarrow{r.m.} X \), then \( X_n \xrightarrow{p} X \).

Proof: Assume \( X_n \xrightarrow{r.m.} X \) for some \( r > 0 \); i.e., \( E(|X_n - X|^r) \to 0 \) as \( n \to \infty \). Then for any \( \epsilon > 0 \),

\[
P(|X_n - X| > \epsilon) = P(|X_n - X|^r > \epsilon^r) \leq \frac{E(|X_n - X|^r)}{\epsilon^r} \to 0 \text{ as } n \to \infty
\]

where the inequality follows from Markov’s inequality applied to the nonnegative random variable \( |X_n - X|^r \).

- Theorem 6

Almost sure convergence implies convergence in probability; i.e., if \( X_n \xrightarrow{a.s.} X \), then \( X_n \xrightarrow{p} X \).

Proof: Assume \( X_n \xrightarrow{a.s.} X \). We want to show that for any \( \epsilon > 0 \) we have \( P(|X_n - X| > \epsilon) \to 0 \) as \( n \to \infty \). Thus, defining the event

\[
A_n(\epsilon) = \{ \omega : |X_n(\omega) - X(\omega)| > \epsilon \}
\]

we want to show that \( P(A_n(\epsilon)) \to 0 \) for any \( \epsilon > 0 \).

Define

\[
A(\epsilon) = \{ \omega : |X_n(\omega) - X(\omega)| > \epsilon \text{ for infinitely many } n \}
\]

Then for any \( \omega \in A(\epsilon) \) we have that \( X_n(\omega) \not\rightarrow X(\omega) \). Since \( X_n \xrightarrow{a.s.} X \) we must have that

\[
P(A(\epsilon)) = 0
\]
Proof cont’d: Now define the event 

$$B_n(\epsilon) = \{ \omega : |X_m(\omega) - X(\omega)| > \epsilon \text{ for some } m \geq n \}$$

Notice that $$B_1(\epsilon) \supset B_2(\epsilon) \supset \cdots \supset B_{n-1}(\epsilon) \supset B_n(\epsilon) \supset \cdots$$; i.e., $$\{B_n(\epsilon)\}$$ is a decreasing sequence of events, satisfying 

$$A(\epsilon) = \bigcap_{n=1}^{\infty} B_n(\epsilon)$$

Therefore by the continuity of probability 

$$P(A(\epsilon)) = P\left(\bigcap_{n=1}^{\infty} B_n(\epsilon)\right) = \lim_{n \to \infty} P(B_n(\epsilon))$$

But since $$P(A(\epsilon)) = 0$$, we obtain that $$\lim_{n \to \infty} P(B_n(\epsilon)) = 0$$.

Example: ($$X_n \overset{p}{\to} X$$ does not imply $$X_n \overset{a.s.}{\to} X$$.) Let $$X_1, X_2, \ldots$$ be independent random variables with distribution 

$$P(X_n = 0) = 1 - \frac{1}{n}, \quad P(X_n = 1) = \frac{1}{n}$$

and show that there is a random variable $$X$$ such that $$X_n \overset{p}{\to} X$$, but 

$$P(\{\omega : X_n(\omega) \to X(\omega)\}) = 0.$$ 

Solution: …

Example: ($$X_n \overset{r.m.}{\to} X$$ does not imply $$X_n \overset{a.s.}{\to} X$$.) Use the previous example…

Proof cont’d: We clearly have 

$$A_n(\epsilon) \subset B_n(\epsilon)$$ for all $$n$$

so $$P(A_n(\epsilon)) \leq P(B_n(\epsilon))$$, so $$P(A_n(\epsilon)) \to 0$$ as $$n \to \infty$$. We obtain 

$$\lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0$$

for any $$\epsilon > 0$$ as claimed. □

Lemma 7 (Sufficient condition for a.s. convergence)

Suppose $$\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$$ for all $$\epsilon > 0$$. Then $$X_n \overset{a.s.}{\to} X$$.

Proof: Recall the event 

$$A(\epsilon) = \{ \omega : |X_n(\omega) - X(\omega)| > \epsilon \text{ for infinitely many } n \}$$

Then 

$$A(\epsilon)^c = \{ \omega : \text{ there exists } N \text{ such that } |X_n(\omega) - X(\omega)| \leq \epsilon \text{ for all } n \geq N \}$$

For $$\epsilon = 1/k$$ we obtain a decreasing sequence of events $$\{A(1/k)^c\}_{k=1}^{\infty}$$. Let 

$$C = \bigcap_{k=1}^{\infty} A(1/k)^c$$

and notice that if $$\omega \in C$$, then $$X_n(\omega) \to X(\omega)$$ as $$n \to \infty$$. Thus if 

$$P(C) = 1$$, then $$X_n \overset{a.s.}{\to} X$$.
We have seen in the proof of Theorem 6 that random variables with marginal distributions given by

\[ P(C) = P\left( \bigcap_{k=1}^{\infty} A(1/k)^c \right) = \lim_{k \to \infty} P(A(1/k)^c) \]

so if \( P(A(1/k)^c) = 1 \) for all \( k \), then \( P(C) = 1 \) and we obtain \( X_n \overset{a.s.}{\to} X \). Thus \( P(A(\epsilon)) = 0 \) for all \( \epsilon > 0 \) implies \( X_n \overset{a.s.}{\to} X \).

As before, let \( A_n(\epsilon) = \{ \omega : |X_n(\omega) - X(\omega)| > \epsilon \} \) and

\[ B_n(\epsilon) = \{ \omega : |X_m(\omega) - X(\omega)| > \epsilon \text{ for some } m \geq n \} \]

We have seen in the proof of Theorem 6 that

\[ P(A(\epsilon)) = \lim_{n \to \infty} P(B_n(\epsilon)) \]

Thus if we can show that \( \lim_{n \to \infty} P(B_n(\epsilon)) = 0 \) for all \( \epsilon > 0 \), then \( X_n \overset{a.s.}{\to} X \).

**Example:** \( (X_n \overset{a.s.}{\to} X \text{ does not imply } X_n \overset{r.m.}{\to} X. \) Let \( X_1, X_2, \ldots \) be random variables with marginal distributions given by

\[ P(X_n = n^3) = \frac{1}{n^2}, \quad P(X_n = 0) = 1 - \frac{1}{n^2} \]

Show that there is a random variable \( X \) such that \( X_n \overset{a.s.}{\to} X \), but \( X_n \overset{r.m.}{\to} X \) if \( r \geq 2/3 \).

**Proof cont’d:** Note that

\[ B_n(\epsilon) = \bigcup_{m=n}^{\infty} A_m(\epsilon) \]

and that the condition \( \sum_{n=1}^{\infty} P(A_n(\epsilon)) = \sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty \) implies

\[ \lim_{n \to \infty} \sum_{m=n}^{\infty} P(A_n(\epsilon)) = 0 \]

We obtain

\[ P(B_n(\epsilon)) = P\left( \bigcup_{m=n}^{\infty} A_m(\epsilon) \right) \]

(by union bound)

\[ \leq \sum_{m=n}^{\infty} P(A_m(\epsilon)) \to 0 \text{ as } n \to \infty \]

so \( \lim_{n \to \infty} P(B_n(\epsilon)) = 0 \) for all \( \epsilon > 0 \). Thus \( X_n \overset{a.s.}{\to} X \). \( \square \)

**Theorem 8**

Convergence in probability implies convergence in distribution; i.e., if \( X_n \overset{P}{\to} X \), then \( X_n \overset{d}{\to} X \).

**Proof:** Let \( F_n \) and \( F \) be the cdfs of \( X_n \) and \( X \), respectively and let \( x \in \mathbb{R} \) be such that \( F \) is continuous at \( x \). We want to show that \( X_n \overset{P}{\to} X \) implies \( F_n(x) \to F(x) \) as \( n \to \infty \).

For any given \( \epsilon > 0 \) we have

\[ F_n(x) = P(X_n \leq x) \]

\[ = P(X_n \leq x, X \leq x + \epsilon) + P(X_n \leq x, X > x + \epsilon) \]

\[ \leq P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon) \]

\[ = F(x + \epsilon) + P(|X_n - X| > \epsilon) \]
Laws of Large Numbers

Let \( X_1, X_2, \ldots \) be a sequence of independent and identically distributed (i.i.d.) random variables with finite mean \( \mu = E(X_i) \) and variance \( \sigma^2 = \text{Var}(X_i) \). The sample mean \( \bar{X}_n \) is defined by

\[
\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

**Theorem 10 (Weak law of large numbers)**

We have \( \bar{X}_n \xrightarrow{p} \mu \); i.e., for all \( \epsilon > 0 \)

\[
\lim_{n \to \infty} P(\{\bar{X}_n - \mu| > \epsilon\}) = 0
\]

Proof: Since the \( X_i \) are independent,

\[
\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{\sigma^2}{n}
\]

Since \( E(\bar{X}_n) = \mu \), Chebyshev’s inequality implies for any \( \epsilon > 0 \)

\[
P(\{\bar{X}_n - \mu| > \epsilon\}) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2 n}
\]

Thus \( P(\{\bar{X}_n - \mu| > \epsilon\}) \to 0 \) as \( n \to \infty \) for any \( \epsilon > 0 \). □
Theorem 11 (Strong law of large numbers)

If $X_1, X_2, \ldots$ is an i.i.d. sequence with finite mean $\mu = E(X_i)$ and variance $\text{Var}(X_i) = \sigma^2$, then $\bar{X}_n \xrightarrow{a.s.} \mu$; i.e.,

$$P\left(\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu\right) = 1$$

Proof: First we show that the subsequence $\bar{X}_{1^2}, \bar{X}_{2^2}, \bar{X}_{3^2}, \ldots$ converges a.s. to $\mu$.

Let $S_n = \sum_{i=1}^{n} X_i$. Then $E(S_n^2) = n^2 \mu$ and $\text{Var}(S_n) = n^2 \sigma^2$. For any $\epsilon > 0$ we have by Chebyshev’s inequality

$$P\left(|\bar{X}_{n^2} - \mu| > \epsilon\right) = P\left(|\frac{1}{n^2} S_{n^2} - \mu| > \epsilon\right) = P\left(|S_{n^2} - n^2 \mu| > n^2 \epsilon\right) \leq \frac{\text{Var}(S_{n^2})}{n^4 \epsilon^2} = \frac{\sigma^2}{n^2 \epsilon^2}$$

Thus $\sum_{n=1}^{\infty} P\left(|\bar{X}_{n^2} - \mu| > \epsilon\right) < \infty$ and Lemma 7 gives $\bar{X}_{n^2} \xrightarrow{a.s.} \mu$.

Proof cont’d: Now we remove the restriction $X_i \geq 0$. Define

$$X_i^+ = \max(X_i, 0), \quad X_i^- = \max(-X_i, 0)$$

and note that $X_i = X_i^+ - X_i^-$ and $X_i^+ \geq 0$, $X_i^- \geq 0$. Letting $\mu^+ = E(X_i^+)$, $\mu^- = E(X_i^-)$ and

$$A_1 = \left\{ \omega : \frac{1}{n} \sum_{i=1}^{n} X_i^+(\omega) \to \mu^+ \right\}, \quad A_2 = \left\{ \omega : \frac{1}{n} \sum_{i=1}^{n} X_i^-(\omega) \to \mu^- \right\}$$

we know that $P(A_1) = P(A_2) = 1$. Thus $P(A_1 \cap A_2) = 1$. But for all $\omega \in A_1 \cap A_2$ we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i^+(\omega) - \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i^-(\omega) = \mu^+ - \mu^- = \mu$$

We conclude that $\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu$. □

Remark: The condition $\text{Var}(X_i) < \infty$ is not needed. The strong law of large numbers (SLLN) holds for any i.i.d. sequence $X_1, X_2, \ldots$ with finite mean $\mu = E(X_1)$.

Example: Simple random walk …

Example: (Single server queue) Customers arrive one-by-one at a service station.

- The time between the arrival of the $(i-1)$th and $i$th customer is $Y_i$, and $Y_1, Y_2, \ldots$ are i.i.d. nonnegative random variables with finite mean $E(Y_1)$ and finite variance.

- The time needed to service the $i$th customer is $U_i$, and $U_1, U_2, \ldots$ are i.i.d. nonnegative random variables with finite mean $E(U_1)$ and finite variance.

Show that if $E(U_1) < E(Y_1)$, then (after the first customer arrives) the queue will eventually become empty with probability 1.
Central Limit Theorem

- If $X_1, X_2, \ldots$ are Bernoulli($p$) random variables, then $S_n = X_1 + \cdots + X_n$ is a Binomial($n, p$) random variable with mean $E(S_n) = np$ and variance $\text{Var}(S_n) = np(1 - p)$.

- Recall the De Moivre-Laplace theorem:
  $$\lim_{n \to \infty} P \left( \frac{S_n - np}{\sqrt{np(1-p)}} \leq x \right) = \Phi(x)$$
  where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt$ is the cdf of a $N(0, 1)$ random variable.

- Since $\Phi(x)$ is continuous at every $x$, the above is equivalent to
  $$\frac{S_n - np}{\sqrt{n}} \xrightarrow{d} X,$$
  where $\mu = E(X_1) = p$ and $\sigma = \sqrt{\text{Var}(X_1)} = p(1 - p)$.

To prove the CLT we will assume that each $X_i$ has MGF $M_{X_i}(t)$ that is defined in an open interval around zero. The key to the proof is the following result which we won’t prove:

**Theorem 13 (Levy continuity theorem for MGF)**

Assume $Z_1, Z_2, \ldots$ are random variables such that the MGF $M_{Z_i}(t)$ is defined for all $t \in (-a, a)$ for some $a > 0$ and all $n = 1, 2, \ldots$ Suppose $X$ is a random variable with MGF $M_X(t)$ defined for $t \in (-a, a)$. If

$$\lim_{n \to \infty} M_{Z_n}(t) = M_X(t) \quad \text{for all } t \in (-a, a)$$

then $Z_n \xrightarrow{d} X$.

The De Moivre-Laplace theorem is a special case of the following general result:

**Theorem 12 (Central Limit Theorem)**

Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean $\mu$ and finite variance $\sigma^2$. Then for $S_n = X_1 + \cdots + X_n$ we have

$$\frac{S_n - np}{\sigma \sqrt{n}} \xrightarrow{d} X,$$

where $X \sim N(0, 1)$.

**Remark:** Note that $\frac{S_n - np}{\sigma \sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$. The SLLN implies that with probability 1, $\bar{X}_n - \mu \to 0$ as $n \to \infty$. The central limit theorem (CLT) tells us something about the speed at which $\bar{X}_n - \mu$ converges to zero.

**Proof of CLT:** Let $Y_n = X_n - \mu$. Then $E(Y_n) = 0$ and $\text{Var}(Y_n) = \sigma^2$. Note that

$$\frac{\sum_{i=1}^{n} Y_i}{\sigma \sqrt{n}} = \frac{S_n - np}{\sigma \sqrt{n}} = Z_n$$

Letting $M_Y(t)$ denote the (common) moment generating function of the $Y_i$, we have by the independence of $Y_1, Y_2, \ldots$

$$M_{Z_n}(t) = M_Y \left( \frac{t}{\sigma \sqrt{n}} \right)^n$$

If $X \sim N(0, 1)$, then $M_X(t) = e^{t^2/2}$. Thus if we show that $M_{Z_n}(t) \to e^{t^2/2}$, or equivalently

$$\lim_{n \to \infty} \ln(M_{Z_n}(t)) = \frac{t^2}{2}$$

then Levy’s continuity theorem implies the CLT.
Proof cont’d: Let \( h = \frac{t}{\sigma \sqrt{n}} \). Then

\[
\lim_{n \to \infty} \ln(M_Z(t)) = \lim_{n \to \infty} n \ln \left( M_Y \left( \frac{t}{\sigma \sqrt{n}} \right) \right) = \frac{t^2}{\sigma^2} \lim_{h \to 0} \frac{M_Y'(h)}{h^2}
\]

It can be shown that \( M_Y(h) \) and all its derivatives are continuous at \( h = 0 \). Since \( M_Y(0) = E(e^{0 \cdot Y}) = 1 \), the above limit is indeterminate. Applying l’Hospital’s rule, we consider the limit

\[
\frac{t^2}{\sigma^2} \lim_{h \to 0} \frac{M_Y'(h)}{2h} = \frac{t^2}{\sigma^2} \lim_{h \to 0} \frac{M_Y'(h)}{2hM_Y(h)}
\]

Again, since \( M'(0) = E(Y) = 0 \), the limit is indeterminate.

Proof cont’d: Apply l’Hospital’s rule again:

\[
\lim_{n \to \infty} \ln(M_Z(t)) = \lim_{h \to 0} \frac{\ln M_Y(h)}{h^2} = \frac{t^2}{\sigma^2} \lim_{h \to 0} \frac{M_Y'(h)}{2hM_Y(h)}
\]

Thus \( M_Z(t) \to e^{t^2/2} \) and therefore \( Z_n = \frac{S_n - n\mu}{\sigma \sqrt{n}} \xrightarrow{d} X \), where \( X \sim N(0, 1) \). \( \Box \)