ECE 6010
Lecture 5 – Sequences and Limit Theorems

Convergent sequences of real numbers and functions

Definition 1 Let $x_1, x_2, \ldots$ be a sequence of real numbers. This sequence converges to a point $x \in \mathbb{R}$ if for every $\epsilon > 0$ there is an $N \in \mathbb{Z}^+$ such that $|x_n - x| < \epsilon$ for all $n \geq$N.

We write $x_n \to x$, or $\lim_{n\to\infty} x_n = x$. □

For real numbers (which are complete), a necessary and sufficient condition:

$$\{x_n\}_{n=1}^{\infty} \text{ converges} \Leftrightarrow \lim_{n\to\infty} \sup_{m>n} |x_m - x_n| = 0.$$

The latter condition says that $\{x_n\}$ is a **Cauchy sequence**.

Definition 2 Suppose $f_1, f_2, \ldots$ is a sequence of *functions* $\Omega \to \mathbb{R}$. This sequence **converges pointwise** to $f : \Omega \to \mathbb{R}$ if $f_n(x) \to f(x)$ for every $x \in \Omega$. That is, for every $x \in \Omega$ and $\epsilon > 0$, there is an $N \in \mathbb{Z}^+$ such that $|f_n(x) - f(x)| < \epsilon$ for all $n \geq N$. □

(It may be necessary to choose a different N for each x.)

Definition 3 We say that f_n converges **uniformly** to f if for each $\epsilon > 0$ there is an $N \in \mathbb{Z}^+$ such that $|f_n(x) - f(x)| < \epsilon$ for all $n \geq N$ **and for all** $x \in \Omega$. □

Modes of convergence of sequences of r.v.s

Suppose $X_1, X_2, \ldots$ is a sequence of random variables defined on $(\Omega, \mathcal{F}, P)$. How can we define a limit of this sequence? As it turns out, there are several different (and inequivalent) ways of defining convergence.

Almost sure convergence

This is a very strong form of convergence, and usually quite difficult to prove.

Definition 4 A sequence of r.v.s $\{X_n\}_{n=1}^{\infty}$ converges **almost surely** (a.s.) to the r.v. X if $P(\Omega_0) = 1$, where

$$\Omega_0 = \{\omega \in \Omega : X_n(\omega) \to X(\omega)\}$$

This is also called convergence **with probability 1.** □

One tool for showing a.s. convergence is the following fact:

$X_n \to X$ a.s. if and only if

$$P(\lim_{n\to\infty} \sup_{m>n} |X_n - X_m| = 0) = 1.$$

Example 1 Let $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}[0, 1]$. Let $X_n(\omega) = ne^{-n\omega}$, $\omega \in [0, 1]$ and $n \in \mathbb{Z}^+$. Note that

$$X_n(\omega) \to 0 \text{ for all } \omega \in (0, 1].$$

$$X_n(0) \to n \text{ (diverges).}$$

So if $P(\{0\}) = 0$ then $X_n \to 0$ a.s. But if $P(\{0\}) > 0$ then X_n doesn't converge in the almost sure sense. □

Mean-square convergence

This is a strong mode of convergence which is usually easier to show than a.s. It is widely used in engineering.

Definition 5 The sequence $\{X_n\}_{n=1}^{\infty}$ converges to the r.v. X in the **mean-square** sense if

$$\lim_{n\to\infty} E[(X_n - X)^2] = 0.$$

□

We write $X_n \to X$ (m.s.) or $X_n \to X$ (q.m.) "quadratic mode."

There is a Cauchy criterion for m.s. convergence: If $E[X_n^2] < \infty$ for all $n \in \mathbb{Z}^+$, then $\{X_n\}$ converges in mean-square if and only if

$$\lim_{n\to\infty} \sup_{m>n} E[(X_m - X_n)^2] = 0.$$

Example 2 Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}[0,1]$, and P is uniform: $P([a,b]) = |b-a|$. Let

$$X_n(\omega) = \begin{cases} n & \omega \in [0, 1/n^3], n \in \mathbb{Z}^+ \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E[X_n^2] = n^2 P([0, 1/n^3]) + 0^2 P([1/n^3, 1]) = n^2 \cdot \frac{1}{n^3} = \frac{1}{n} \to 0.$$

So $X_n \to 0$ m.s.

What about a.s. convergence in this case? □

Here is an interesting fact: If $X_n \to X$ (m.s.) and $X_n \to Y$ (a.s.), then $X = Y$ (a.s.).

Convergence in Probability

Definition 6 The sequence $\{X_n\}_{n=1}^{\infty}$ converges to X **in probability** (i.p.) if

$$P(|X_n - X| > \epsilon) \to 0$$

as $n \to \infty$ for every $\epsilon > 0$.

Equivalently, we say that

$$P(|X_n - X| \leq \epsilon) \to 1.$$

□

Example 3 Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}[0,1]$, and P is uniform. Let

$$X_n = \begin{cases} n & \omega \in [0, 1/n] \\ 0 & \text{otherwise.} \end{cases}$$

Note: $X_n \to 0$ (a.s.), but X_n does not converge in m.s.

$$P(|X_n - 0| > \epsilon) = P([0, 1/n]) = 1/n \to 0,$$

so $X_n \to 0$ (i.p.) □

Convergence in Distribution

Definition 7 The sequence $\{X_n\}_{n=1}^{\infty}$ converges **in distribution** (or **in law**) to the random variable X if $F_{X_n}(x) \to F_X(x)$ at all continuous points of F_X. □

Example 4 $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}[0,1]$, P is uniform. Let

$$X_n = \begin{cases} 1 & \omega \geq 1/n \\ 0 & \omega < 1/n. \end{cases}$$

Then

$$F_{X_n}(x) = (1/n)u(x) + (1 - 1/n)u(x - 1)$$

Then $X_n \to X$, where $F_X(x) = u(x - 1)$. (Draw the distributions.) $\square$

Note from this example that the X_n values don't really "approach" any value — the values are still 1 and 0. This is in distinction to the first three modes of convergence, in which $|X_n - X| \to 0$ in some sense.

By the definition of this mode of convergence, we don't have to worry about the points of discontinuity of F_{X_n}.

Example 5 Let $X_n = 1/n$ for all $\omega \in \Omega$. (So it doesn't matter what the underlying P is.) The pointwise convergence is

$$\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 1 & x > 0 \qquad \text{(not a c.d.f.)} \\ 0 & x \leq 0. \end{cases}$$

(Why isn't this a c.d.f?)

Take

$$F_X(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0. \end{cases}$$

This is a c.d.f, but different from $\lim F_{X_n}$. However the difference is at a point of discontinuity.

Hence $X_n \to 0$ (in distribution). $\square$

Why and Which?

We have defined several different modes of convergence. Why so many? The basic answer is that they are inequivalent — one does not meet all the analytical needs. Some are stronger than others.

$$X_n \to X \text{ (m.s.)} \Rightarrow X_n \to X \text{ (i.p.)}$$

$$X_n \to X \text{ (a.s.)} \Rightarrow X_n \to X \text{ (i.p.)}$$

$$X_n \to X \text{ (i.p..)} \Rightarrow X_n \to X \text{ (in distribution)}$$

So convergence in distribution is weaker than i.p., m.s. or a.s.

In general, none of the implications can be reversed. And m.s. and a.s. do not imply each other. (Venn diagram – dist. on the outside, then i.p., with m.s. and a.s. overlapping inside.)

Proof of $X_n \to X$ (m.s.) $\Rightarrow X_n \to X$ (i.p.)

By Chebyshev's inequality, for every $\epsilon > 0$, $P(|X_n - X| > \epsilon) \leq E[(X_n - X)^2]/\epsilon^2$. So if $E[(X_n - X)^2] \to 0$, then $P(|X_n - X| > \epsilon) \to 0$ for all ϵ. $\square$

Proof of $X_n \to X$ (a.s.) $\Rightarrow X_n \to X$ (i.p.)

Choose $\epsilon > 0$. Write $B_n = \{\omega \in \Omega | \sup_{m\geq n} |X_m(\omega) - X(\omega)| > \epsilon\}$. This is a decreasing family of sets: Suppose $n_1 < n_2$ and $\omega \in B_{n_2}$. Then by the definition, $\omega \in B_{n_1}$, so $B_{n_2} \subset B_{n_1}$.

Note that $\cap_{n=1}^{\infty} B_n = \lim_{n\to\infty} B_n$.

Now consider the set $Z = \{\omega \in \infty : X_n(\omega) \not\to X(\omega)\}$. For a given ϵ we see that $\cap_{n=1}^{\infty} B_n$ is a subset of Z. Since $X_n \to X$ a.s., we have $P(X_n \not\to X) = 0$. Thus

$$P(\lim_n B_n) \leq P(X_n \not\to X) = 0$$

Now notice that $\{\omega : |X_n - X| > \epsilon\} \in B_n$ (since we are looking at only one point, and not $\sup_{m\geq n}$. So $P(|X_n - X| > \epsilon) \leq P(B_n)$, which we just showed $\to 0$.

$\square$

Proof of $X_n \to X$ (i.p.) $\Rightarrow X_n \to X$ (in distribution)

Suppose $X_n \to X$ (i.p.). Choose $\epsilon > 0$ and let x be a continuity point of F_X. Then

$$F_X(x-\epsilon) = P(X \leq x-\epsilon) = \underbrace{P(X \leq x-\epsilon, X_n \leq x)} + P(X \leq x-\epsilon, X_n > x).$$

$$F_{X_n}(x) = \underbrace{P(X \leq x-\epsilon, X_n \leq x)} + P(X > x-\epsilon, X_n \leq x).$$

Solving for the bracketed term in the second and substituting it into the first we obtain

$$\begin{aligned} F_X(x-\epsilon) &= F_{X_n}(x) + P(X \leq x-\epsilon, X_n > x) - P(X > x-\epsilon, X_n \leq x) \\ &\leq F_{X_n}(x) + P(X \leq x-\epsilon, X_n > x). \end{aligned}$$

Observe that

$$\{X \leq x-\epsilon, X_n > x\} \subset \{|X_n - X| > \epsilon\}$$

(for example, let $X_n = x + \delta_1$ and $X = x - \epsilon - \delta_2$ with $\delta_1 > 0$ and $\delta_2 \geq 0$. Then $|X_n - X| = |\epsilon + \delta_1 + \delta_2| > \epsilon$) so that $P(\{X \leq x-\epsilon, X_n > x\}) \leq P(\{|X_n - X| > \epsilon\})$. Thus

$$F_X(x-\epsilon) \leq F_{X_n}(x) + P(|X_n - x| > \epsilon).$$

Similarly,

$$F_{X_n}(x) \leq F_X(x+\epsilon) + P(|X_n - X| > \epsilon).$$

Since we have convergence in probability, $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$, so

$$F_X(x-\epsilon) \leq \lim_{n\to\infty} F_{X_n}(x)$$

Similarly also

$$F_X(x+\epsilon) \geq \lim_{n\to\infty} F_{X_n}(x).$$

Combining these,

$$F_X(x-\epsilon) \leq \lim_{n\to\infty} F_{X_n}(x) \leq F_X(x+\epsilon).$$

Since x is a continuity point of F_X and ϵ is chosen arbitrarily, we can write

$$\lim_{\epsilon\to 0} F_X(x-\epsilon) = F_X(x) \leq \lim_{n\to\infty} F_{X_n}(x) \leq F_X(x) = \lim_{\epsilon\to 0} F_X(x+\epsilon).$$

So $F_{X_n}(x) \to F_X(x)$. (Convergence in distribution.) □

Some examples of invalid implications

To see which modes are "stronger" than others, we can consider some counterexamples.

Example 6 Let $X_n \to X$ (i.p.) Can we say that $X_n \to X$ (m.s.)?

Let $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}[0,1], \text{uniform})$. Let

$$X_n = \begin{cases} n & \omega \in [0, 1/n] \\ 0 & \text{otherwise.} \end{cases}$$

We've shown that $X_n \to 0$ (i.p.), but $E[X_n^2] \to \infty$ so $X_n \not\to 0$ (m.s.).

Since $X_n \to 0$ (a.s.), we also see that a.s. $\not\Rightarrow$m.s. □

Example 7 Does i.p. imply a.s.? Define a sequence of r.v.s as follows on $\Omega = [0,1]$: $X_1(\omega) = 1$. For X_2, X_3, divide into two parts $[0,1/2), (1/2,1]$, with $X_2(\omega) = 1$ on the first half, and $X_3(\omega) = 1$ on the second half. For X_4, X_5, X_6, X_7 split into fourths, with $X_4(\omega) = 1$ on the first fourth, etc.

$$P(|X_n - 0| > \epsilon) = P(X_n = 1)$$

which decreases (at a rate approximately $1/\log n$) as $n \to \infty$. So $X_n \to 0$ (i.p.)

However, for a.s. convergence, we see that X_n alternates (non uniformly) between 0 and 1. So $X_n \not\to 0$ (a.s.)

Note that this example also converges in m.s. because the 2nd moment is $P(X_n = 1) \approx 1/\log_2(n) \to 0$. $\square$

Example 8 What about convergence in distribution and convergence i.p.?

Let $X \sim \mathcal{N}(0, 1)$, and $X_n = (-1)^n X$. Note that $X_n \sim \mathcal{N}(0, 1)$. So $F_{X_n} = F_x$ for all n. But

$$P(|X_n - X| > \epsilon) = \begin{cases} P(|X| > \epsilon/2) & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

So it does not $\to 0$ for all ϵ; it alternates.

All the other modes of convergence depend on joint distributions, but convergence in distribution depends on marginals, which don't tell us the whole picture. $\square$

Some other relationships:

1. If $X_n \to X$ (i.p.) then there is a subsequence $\{X_{n_k}\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} X_{n_k} = X$ (a.s.)
2. If $X_n \to X$ and there is a r.v. Y with finite second moment such that $|X_n| \le Y$ (a.s.) for every $n \in \mathbb{Z}^+$, then $X_n \to X$ (m.s.).
3. If $X_n \to C$ (in distribution), then $X_n \to C$ (i.p.)

Limit Theorems

Laws of Large Numbers

Suppose $X_1, X_2, \ldots,$ is a sequence of r.v.s. We are often interested in sums $\sum_{i=1}^n X_i$, as n becomes large. What can we say about such sums?

Suppose all X_i have the same means μ, $E[X_i] = \mu$, and are uncorrelated. We would expect the **average** $\frac{1}{n}\sum_{i=1}^n X_i$ to "approach" μ in some way as $n \to \infty$. If $\text{var}(x_i) < \infty$, consider

$$\frac{1}{n}\sum_{i=1}^n X_i - \mu.$$

Let us look at m.s. convergence;

$$\begin{aligned} E\left[\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)^2\right] &= E\left[\left(\frac{1}{n}\sum_i (X_i - \mu)\right)^2\right] \\ &= \frac{1}{n^2} E[\sum_i (X_i - \mu)^2] \\ &= \frac{1}{n^2} E[\sum_i \sum_j (X_i - \mu)(X_j - \mu)] \\ &= \frac{1}{n^2}\sum_i\sum_j \text{cov}(X_i, X_j) = \frac{1}{n^2}\sum_{i=1}^n \text{cov}(X_i, X_j) \\ &= \frac{1}{n^2}\sum_{i=1}^n \text{var}(X_i) \end{aligned}$$

Summarizing: If $E[X_i] = \mu$ and $\{X_i\}$ are mutually uncorrelated and have finite variance, $\frac{1}{n}\sum_{i=1}^n \text{var}(X_i) \to 0$, so that

$$\frac{1}{n}\sum_{i=1}^n X_i \to \mu(\text{m.s.}) \Rightarrow \frac{1}{n}\sum_{i=1}^n X_i \to \mu(\text{i.p.}).$$

This is an example of a **weak law of large numbers.**

Definition 8 Suppose $\{X_i\}_{i=1}^\infty$ is a sequence of r.v.s and $\{b_i\}_{i=1}^\infty$ is a sequence of reals diverging to ∞. Then $\{X_i\}_{i=1}^\infty$ satisfies a **weak law of large numbers** (WLLN) if there is another sequence $\{a_i\}_{i=1}^\infty$ of real numbers such that

$$\frac{1}{b_n}\sum_{i=1}^n X_i - a_i \to 0(\text{i.p.}).$$

□

In the example we just gave, $b_n = n$ and $a_i = \mu$.

Definition 9 A **strong law** of large numbers is the same as the preceding definition, except that convergence is almost sure (a.s.). □

Kolmogorov's Strong Law

Definition 10 An infinite sequence of r.v.s is independent if every finite subcollection of the r.v.s is independent. □

Theorem 1
(Kolmogorov's Strong Law) Suppose $\{X_n\}_{n=1}^\infty$ is a sequence of independent r.v.s with finite means for each i. If

$$\sum_{i=1}^n \frac{\text{var}(X_i)}{b_i^2} < \infty$$

then

$$\frac{1}{b_n}\sum_{i=1}^n X_i - a_n \to 0(a.s.),$$

where

$$a_n = \frac{\sum_{i=1}^n \mu_i}{b_n}$$

Example 9 If $b_n = n$ and $\mu_n = \mu$, then Kolmogorov's law implies:

$$\sum_{i=1}^n \frac{\text{var}(X_i)}{i^2} < \infty \Rightarrow \frac{1}{n}\sum_{i=1}^n X_i \to \mu \text{ (a.s.)}.$$

□

Note that in the case that all the variances are bounded, e.g.

$$\text{var}(X_i) < \sigma^2 < \infty \text{ for all } i$$

then

$$\sum_{i=1}^\infty \frac{\text{var}(X_i)}{i^2} \le \sigma^2 \sum_{i=1}^\infty \frac{1}{i^2} < \infty$$

So, if the variances grow *sublinearly*, the theorem can apply.

We can get an even stronger conclusion:

Theorem 2
Kinchine's Strong Law of Large Numbers.
Suppose $\{X_i\}_{i=1}^{\infty}$ is an i.i.d. sequence (i.e., a sequence of i.i.d. r.v.s) with finite mean

$$|E[X_i]| = |\mu| < \infty.$$

Then the sample mean converges almost surely to the ensemble mean:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \text{ (a.s.)}$$

Proving these types of theorems

The proofs follow from more general limit theorems.

Definition 11 Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events. The **limit superior** (lim sup) of $\{A_n\}$ is

$$\limsup_n A_n = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k$$

This is the set of all points that are in $\{A_n\}$ **infinitely often.** □

So $\omega \in \limsup_n A_n \Leftrightarrow \omega$ is in infinitely many of the sets A_n. (It keeps coming back.)
Another notation is: $\limsup_n A_n = A_n$ i.o. (infinitely often).
We observe that if $A_n \uparrow A$ or $A_n \downarrow A$ then A_n (i.o.) $= A$.

Lemma 1
The **Borel Cantelli lemma**. *[This is frequently a good problem for math qualifiers.]*

1. *If $\sum_{i=1}^{\infty} P(A_n) < \infty$ then $P(A_n \text{ (i.o.)}) = 0$. That is, $P(A_n) \to 0$.*
2. *(Conversely) If $\{A_n\}_{n=1}^{\infty}$ are independent events and $\sum_{i=1}^{\infty} P(A_n) = \infty$ then $P(A \text{ (i.o.)}) = 1$.*

Proof

1. $\underbrace{\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k}_{A_n \text{ (i.o.)}} \subset \cup_{k=n}^{\infty} A_k$ for all n. So

$$P(A_n \text{ (i.o.)}) \le P(\cup_{k=n}^{\infty} A_k) \le \sum_{k=n}^{\infty} P(A_k) \to 0$$

as $n \to \infty$ if $\sum_{k=1}^{\infty} P(A_k) < \infty$. So $P(A_n \text{ (i.o.)}) = 0$ if $\sum_{k=1}^{\infty} P(A_k) < \infty$.

2. Using DeMorgan's law,

$$[A_n \text{ (i.o.)}]^c = \cup_{k=1}^{\infty} \cap_{k=n}^{\infty} A_k^c.$$

Pick n and N with $n < N$. Consider

$$\begin{aligned} P(\cap_{k=n}^{\infty} A_k^c) &= \prod_{k=n}^{N} P(A_k^c) \text{ (by independence)} \\ &= \prod_{k=n}^{N} 1 - P(A_k) \le \prod_{k=n}^{N} e^{-P(A_k)} \qquad \text{since } 1 - x \le e^{-x} \\ &= \exp[-\sum_{k=n}^{N} P(A_k)]. \end{aligned}$$

If $\sum_{k=1}^{\infty} P(A_k)$ diverges, then $\sum_{k=n}^{\infty} P(A_k)$ diverges too, and thus

$$\lim_{N\to\infty} \exp[-\sum_{k=n}^{N} P(A_k)] \to 0.$$

So
$$\lim_{N\to\infty} P(\cap_{k=n}^{N} A_k^c) = 0$$
for all n, i.e.,
$$P(\cap_{k=n}^{\infty} A_k^c) = 0$$
for all n.

Now, $\limsup_n A_n$ is just the union of all of those intersections, so
$$P(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k^c) \le \sum_{n=1}^{\infty} P(\cap_{k=n}^{\infty} A_k^c) = 0.$$
so that $P(A_n \text{ (i.o.)}) = 1$.

$\square$

Kolmogorov's Inequality

Suppose $X_1, X_2, \ldots,$ are independent with zero means and finite variances. Define S_n to be the running sum
$$S_n = \sum_{k=1}^{n} X_k$$
Then for each $\alpha > 0$,
$$P(\max_{1\le k\le n} |S_k| \ge \alpha) \le \frac{1}{\alpha^2}\operatorname{var}(S_n).$$
This is a lot like the Chebyshev inequality, but instead of looking at the variance of all of the terms, we simply look at the variance of the last one.

Central Limit Theorems

Theorem 3
Central Limit Theorem *Suppose $\{X_n\}$ is a sequence of i.i.d. random variables with mean $mu < \infty$ and variance $\sigma^2 < \infty$. Then*
$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \to X \textit{ (in distribution)}$$
where
$$X \sim \mathcal{N}(0, \sigma^2).$$
That is,
$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \le x\right) \to \int_{-\infty}^{x} e^{-t^2/2\sigma^2}\, dt$$
The main point: Sums of i.i.d. random variables tend to look **Gaussian**.

To work our way up to this, here are a couple of lemmas:

Lemma 2
Suppose $\{X_n\}$ is a sequence of r.v.s with characteristic functions $\{\phi_n\}$. If there exists a r.v. X with ch.f. ϕ such that
$$\lim_{n\to\infty} \phi_n(u) = \phi(u)$$
for all $u \in \mathbb{R}$ then
$$X_n \to X \textit{ (in distribution)}.$$

Lemma 3
Suppose X is a r.v. with $E[X^2] < \infty$. Then ϕ_X has the expansion

$$\phi_X(u) = 1 + iuE[X] - \frac{u^2}{2}(E[X^2] + \delta(u))$$

where $\lim_{u\to 0} \delta(u) = 0$.

Proof of the Central Limit Theorem. For convenience (w.o.l.o.g.), take $\mu = 0$. Define $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$. Then

$$\begin{aligned}
\phi_{S_n}(u) &= E[\exp(iuS_n)] = E[\exp(iu/\sqrt{n} \sum_i X_i)] = \prod_{i=1}^n E[\exp[iu/\sqrt{n} X_i]] \\
&= [\phi_X(u/\sqrt{n})]^n \\
&= [1 + iu/\sqrt{n}\mu - (u/\sqrt{n})^2/2(E[X_i^2] + \delta(u/\sqrt{n}))]^n \\
&= [1 - u^2/2n(\sigma^2 + \delta(u/\sqrt{n}))]^n
\end{aligned}$$

From "elementary" calculus we recall that

$$(1 + a_n)^n \to e^{\lim na_n}$$

Thus

$$\phi_{S_n}(u) \to \exp\left[\lim_{n\to\infty}(-u^2/2(\sigma^2 + \delta(u/\sqrt{n})))\right] = \exp(-\sigma^2 u^2/2).$$

This is the form of a characteristic function of a Gaussian (with zero mean). □

Summarizing, if X_k ahas zero mean and variance 1,

$$\frac{1}{n}\sum_{k=1}^n X_k \to 0 \text{ (a.s.)}$$

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k \to \mathcal{N}(0, 1) \text{ (in distribution)}$$