Lecture 7: Convergence and Limit Theorems

- Almost sure convergence
- Convergence in mean square
- Convergence in probability
- Convergence in distribution

Motivation: Estimating Mean

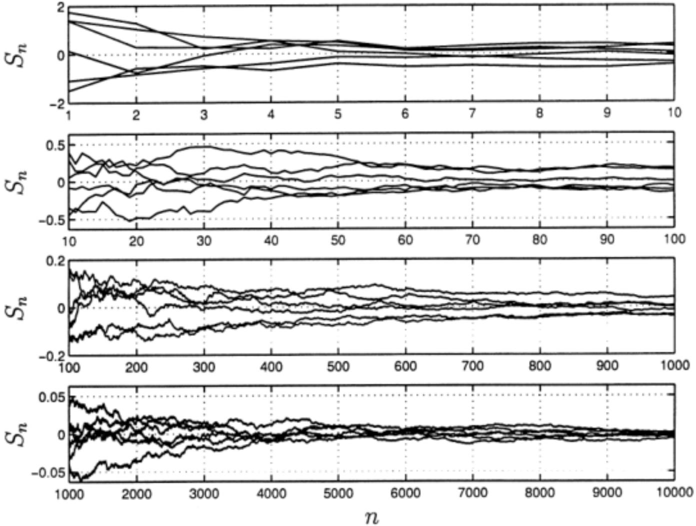
- Want to estimate the mean $\mathbb{E}[X]$ of a distribution.
- Generate $X_1, X_2, \ldots, X_n$ i.i.d. drawn from the distribution and compute the sample mean

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

- Does S_n converge to $\mathbb{E}[X]$? In what sense? Note that S_n is a random variable.

Motivation: Estimating Mean

- Example: $X_1, X_2, \ldots, X_n$ i.i.d. $\mathcal{N}(0, 1)$
  - Generate 6 sets of $X_1, X_2, \ldots, X_n$ (sample paths).
  - Note that every sample path appears to be converging to the mean 0 as n increases.



Almost Sure Convergence

- Recall that a sequence of numbers $x_1, x_2, \ldots$ converges to x if given any $\epsilon > 0$, there exists an $n(\epsilon)$ such that for all $n \geq n(\epsilon)$, $|x_n - x| < \epsilon$. Equivalently, $\max_{i \geq n} |x_i - x| \to 0$ as $n \to \infty$.

- Suppose a sequence of random variables $X_1, X_2, \ldots$ defined over the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For every $\omega \in \Omega$, we get a sample path $X_1(\omega), X_2(\omega), \ldots$, which is a sequence of numbers.

- We say $X_n \to X$ almost surely (a.s.) or with probability 1 (w.p. 1) if

$$\mathbb{P}\Big(\Big\{\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\Big\}\Big) = 1.$$

- This means that the set of ω such that the sample path or sequence $X_1(\omega), X_2(\omega), \ldots$ converges to $X(\omega)$ has probability 1.

Almost Sure Convergence

Lemma. $\mathbb{P}\left(\max_{i\geq n}|X_i - X| > \epsilon\right) \to 0$ for any $\epsilon > 0$ iff $X_n \xrightarrow{\text{a.s.}} X$.

We prove the forward direction. We will see that the converse holds when discussing convergence in probability.

Let $M_n = \max_{i\geq n}|X_i - X|$, which is a decreasing sequence bounded below by 0. Therefore $M_n \downarrow M$ for some M.

Then, for all $\epsilon > 0$, $\mathbb{P}(M > \epsilon) \leq \mathbb{P}(M_n > \epsilon) \to 0$ as $n \to \infty$.

Therefore, from continuity of $\mathbb{P}$, $\mathbb{P}(M = 0) = 1$ and $M_n \to 0$ a.s. This is equivalent to saying that $X_n \to X$ a.s.

- Example †: $X_1, X_2, \ldots$ i.i.d. $\sim \mathrm{Bern}\,(1/2)$. Let $Y_n = 2^n \prod_{i=1}^n X_i$. Show that the sequence Y_n converges to 0 a.s.
  For $0 < \epsilon < 1$, we have

$$
\begin{aligned}
\mathbb{P}\left(\max_{i \geq n} |Y_i - 0| > \epsilon \right) &\leq \mathbb{P}(X_i = 1 \; \forall \, i \leq n) \\
&= \frac{1}{2^n} \to 0
\end{aligned}
$$

  as $n \to \infty$.
- Important example of a.s. convergence is the Strong Law of Large Numbers (SLLN): If $X_1, X_2, \ldots$ are i.i.d. (in fact *pairwise* independence suffices) with finite mean μ, then

$$
S_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{a.s.}} \mu
$$

  as $n \to \infty$. Proof of the SLLN is beyond the scope of this course. See [Github: https://github.com/wptay/aipt].

Convergence in Mean Square

- We say $X_n \to X$ in mean square or L^2 if

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

- Let $X_1, X_2, \ldots$ be i.i.d. with finite mean $\mathbb{E}[X]$ and $\mathrm{var}(X)$. Then $S_n \to \mathbb{E}[X]$ in mean square.

$$\begin{aligned}
\mathbb{E}\left[(S_n - \mathbb{E}[X])^2\right] &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X])\right)^2\right] \\
&= \frac{1}{n^2}\mathbb{E}\left[\sum_{i,j}(X_i - \mathbb{E}[X])(X_j - \mathbb{E}[X])\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}(X_i - \mathbb{E}[X])^2 \quad \because\ X_i\text{s are independent} \\
&= \frac{1}{n}\,\mathrm{var}(X) \to 0
\end{aligned}$$

as $n \to \infty$.

- Note proof works even if the X_is are only pairwise independent or even only uncorrelated.

Convergence in Mean Square

- Convergence in mean square does not imply convergence a.s.
- Example ♣: Let $X_1, X_2, \ldots$ be independent random variables such that

$$X_n = \begin{cases} 0 & \text{w.p. } 1 - \frac{1}{n}, \\ 1 & \text{w.p. } \frac{1}{n}. \end{cases}$$

- $\mathbb{E}\left[(X_n - 0)^2\right] = \frac{1}{n} \to 0$ as $n \to \infty$. Therefore, this sequence converges to 0 in mean square.
- For any $0 < \epsilon < 1$, using the inequalities $\log(1-x) \le -x$ and $\sum_{i=n}^{m} \frac{1}{i} \ge \log \frac{m}{n}$,

$$\begin{aligned} \mathbb{P}\left(\max_{i \ge n} |X_i| < \epsilon\right) &= \lim_{m\to\infty} \mathbb{P}\left(\max_{n \le i \le m} |X_i| < \epsilon\right) = \lim_{m\to\infty} \prod_{i=n}^{m}\left(1 - \frac{1}{i}\right) \\ &= \lim_{m\to\infty} \exp\left(\sum_{i=n}^{m} \log\left(1 - \frac{1}{i}\right)\right) \le \lim_{m\to\infty} \exp\left(-\sum_{i=n}^{m} \frac{1}{i}\right) \le \lim_{m\to\infty} \frac{n}{m} = 0. \end{aligned}$$

Therefore, it does not converge a.s.

Convergence in Mean Square

- Convergence a.s. does not imply convergence in mean square.

- From Example †, Y_n converges to 0 a.s., but

$$\mathbb{E}\big[(Y_n - 0)^2\big] = 2^{2n}\frac{1}{2^n} = 2^n,$$

  the sequence does not converge in mean square.

Convergence in Probability

- A sequence $X_1, X_2, \ldots$ converges to a random variable X in probability if for any $\epsilon > 0$,

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

- If $X_n \to X$ a.s., then $X_n \xrightarrow{\text{p}} X$.
  For proof, see [Github: https://github.com/wptay/aipt].
  If $X_n \to X$ a.s., then $\mathbf{1}_{\{|X_n - X| > \epsilon\}} \to 0$ a.s. The Dominated Convergence Theorem tells us that

$$\mathbb{E}\mathbf{1}_{\{|X_n - X| > \epsilon\}} = \mathbb{P}(|X_n - X| > \epsilon) \to 0.$$

- Example ♣:

$$\mathbb{P}(|X_n - 0| > \epsilon) = \frac{1}{n} \to 0,$$

so the sequence converges in probability. But we saw before that it does not converge a.s.

Convergence in Probability

- If $X_n \to X$ in mean square, then $X_n \xrightarrow{\mathrm{p}} X$.

  From the Markov inequality,

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{1}{\epsilon^2}\mathbb{E}(X_n - X)^2 \to 0.$$

- Converse is not true. Example:

$$X_n = \begin{cases} 0 & \text{w.p. } 1 - \frac{1}{n}, \\ n & \text{w.p. } \frac{1}{n}. \end{cases}$$

  Convergence in probability:

$$\mathbb{P}(|X_n - 0| > \epsilon) = \frac{1}{n} \to 0.$$

  But,

$$\mathbb{E}(X_n - 0)^2 = n^2 \cdot \frac{1}{n} = n \to \infty.$$

- Convergence in probability is weaker than both convergence a.s. and in mean square.

Weak Law of Large Numbers

Suppose $X_1, X_2, \ldots$ are such that $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = \sigma^2 < \infty$ and $\mathbb{E}[X_i X_j] \leq 0$ for $i \neq j$. Then,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{p}} 0 \text{ as } n \to \infty.$$

Proof:

We have

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] &= \frac{1}{n^2}\mathbb{E}\left[\sum_{i}^{n} X_i^2 + 2\sum_{i<j} X_i X_j\right] \\
&\leq \frac{1}{n^2}\sum_{i}^{n} \mathbb{E}X_i^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

Weak Law of Large Numbers

From Chebyshev's inequality, we then have for any $\epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}\frac{\sigma^2}{n} \to 0,$$

as $n \to \infty$.

Convergence in distribution

- We say that $X_n \xrightarrow{\mathrm{d}} X$ if for all continuous bounded functions f,

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)].$$

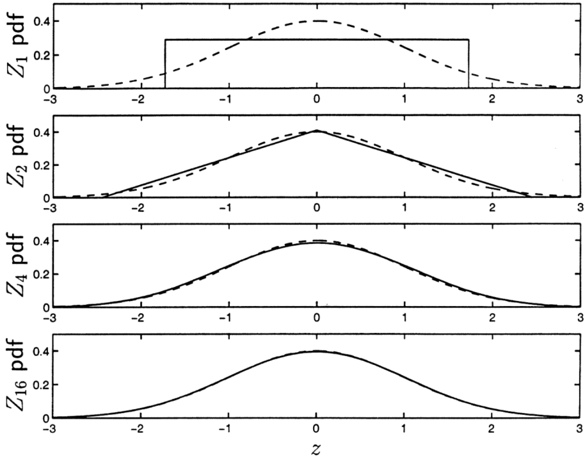
- Equivalent to saying $X_n \xrightarrow{\mathrm{d}} X \iff F_{X_n}(t) \to F_X(t)$ for all continuity points t of $F_X(\cdot)$. [Github: https://github.com/wptay/aipt]

- Central Limit Theorem (CLT) for i.i.d. sequences.
  Consider i.i.d. random variables $X_1, X_2, \ldots$, with $\mathbb{E}X_i = \mu$ and $\operatorname{var} X_i = \sigma^2 < \infty$. Let $S_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then

$$Z_n = \frac{S_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

CLT Example

- Example: Let $X_1, X_2, \ldots$ be i.i.d. $\mathrm{U}[-1, 1]$ r.v.s and Z_n be as defined before. The following plots show the pdf of Z_n for $n = 1, 2, 4, 16$. Note how fast the pdf of Z_n becomes close to gaussian



Characteristic Functions

- Given a random variable X, the characteristic function of X is $\varphi : \mathbb{R} \mapsto \mathbb{C}$ given by

$$\begin{aligned}\varphi(t) &= \mathbb{E}\left[e^{\mathfrak{i}tX}\right], \\ &= \mathbb{E}[\cos(tX)] + \mathfrak{i}\mathbb{E}[\sin(tX)],\end{aligned}$$

  where $\mathfrak{i} = \sqrt{-1}$.
- This is the Fourier transform of pdf of X.
- Example: The characteristic function of the Gaussian distribution $N(\mu, \sigma^2)$ is

$$\varphi(t) = e^{\mathfrak{i}\mu t - \frac{\sigma^2 t^2}{2}}.$$

  In particular, when $\mu = 0, \sigma^2 = 1$,

$$\varphi(t) = e^{-\frac{t^2}{2}}.$$

Fourier Inversion

- Suppose that $\varphi(t)$ is a characteristic function of a r.v. X. If $\int |\varphi(t)| \,\mathrm{d}t < \infty$, then X has pdf

$$f_X(x) = \frac{1}{2\pi} \int \varphi(t) e^{-\mathrm{i}tx} \,\mathrm{d}t.$$

- If $X_n \xrightarrow{\mathrm{d}} X$ then $\varphi_n(t) \to \varphi(t),\ \forall\, t \in \mathbb{R}$.
  Obvious because $\varphi_n(t) = \mathbb{E}\left[e^{\mathrm{i}tX_n}\right] = \mathbb{E}[\cos tX_n] + \mathrm{i}\mathbb{E}[\sin tX_n]$, $\cos(tx)$ and $\sin(tx)$ are bounded continuous functions, and $X_n \xrightarrow{\mathrm{d}} X$.

Fourier Inversion

- Converse not true. The sequence $X_n \sim \mathcal{N}(0, n)$ has characteristic function $\varphi_n(t) = e^{-\frac{nt^2}{2}} \xrightarrow{n\to\infty} 0,\ \forall\, t \neq 0$ and $\varphi_n(0) = 1,\ \forall\, n$. Therefore, $\varphi_n(t)$ converges. But for all $x \in \mathbb{R}$,

$$\mathbb{P}(X_n < x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2n}}\right)\right) \to \frac{1}{2},$$

when $n \to \infty$, which implies that this sequence does not converge in distribution.

Levy's Continuity Theorem

- Suppose X_n has characteristic function $\varphi_n(t) \to \varphi(t)$, and $\varphi(t)$ is continuous at $t = 0$. Then there exists X s.t. $\varphi(t)$ is the characteristic function of X and $X_n \xrightarrow{\mathrm{d}} X$.
  For proof, see [Github: https://github.com/wptay/aipt].
- Proof of CLT for i.i.d. sequence. Without loss of generality, we assume that $\mu = 0, \sigma = 1$. We have

$$Z_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j.$$

  Let

$$\begin{aligned}
\varphi_n(t) &= \mathbb{E}\left[\exp\left(\mathrm{i}t \frac{\sum_{j=1}^{n} X_j}{\sqrt{n}}\right)\right] \\
&= \prod_{j=1}^{n} \mathbb{E}\left[\exp\left(\mathrm{i}t \frac{X_j}{\sqrt{n}}\right)\right] \\
&= \left(\mathbb{E}e^{\mathrm{i}t\frac{X_1}{\sqrt{n}}}\right)^n \\
&= \varphi_1\left(\frac{t}{\sqrt{n}}\right)^n.
\end{aligned}$$

CLT Proof

- From the Taylor series expansion, we have

$$\varphi_1(t) = \varphi_1(0) + \varphi_1'(0)t + \frac{\varphi_1''(0)}{2!} + o(t^2).$$

Since $\varphi_1(0) = 1$, $\varphi_1'(0) = \mathbb{E}\left[\mathrm{i}X_1 e^{i0X_1}\right] = \mathrm{i}\mathbb{E}X_1 = 0$, $\varphi_1''(0) = \mathbb{E}\left[(\mathrm{i}X_1)^2\right] = -1$, we obtain

$$\varphi_1(t) = 1 - \frac{t^2}{2} + o(t^2).$$

Therefore,

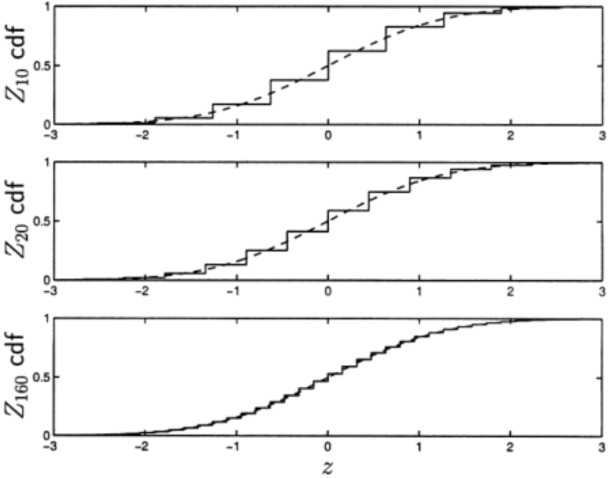
$$\begin{aligned}\varphi_n(t) &= \varphi_1\left(\frac{t}{\sqrt{n}}\right)^n \\ &= \left(1 - \frac{t^2}{2n} + o(t^2/n)\right)^n \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}},\end{aligned}$$

the characteristic function of $\mathcal{N}(0,1)$ (certainly continuous at $t=0$). From Levy's Continuity Theorem, we then have

$$Z_n \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

CLT for Discrete R.V.

Example: $X_1, X_2, \dots$ are i.i.d. $\sim \mathrm{Bern}\,(1/2)$. Note that $Z_n = \sum_{i=1}^{n} \frac{X_i - 1/2}{\sqrt{n}/2}$ is discrete and has no pdf. But its cdf converges to the Gaussian cdf.



CLT for Random Vectors

- The CLT applies to i.i.d. sequences of random vectors.
- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be i.i.d. with each having finite mean $\boldsymbol{\mu}$ and non-singular covariance matrix $\mathbf{\Sigma}$.

$$\mathbf{Z}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{X}_i - \boldsymbol{\mu}) \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$$

Summary

$$\begin{tikzcd}
\text{with probability 1} \arrow[dr, Rightarrow] & & \\
& \text{in probability} \arrow[r, Rightarrow] & \text{in distribution} \\
\text{in mean square} \arrow[ur, Rightarrow] & &
\end{tikzcd}$$