Recap: Convergence in Probability

- A sequence of random variables, X_n, is said to **converge in probability** to a random variable X_0 is

$$\lim_{n \to \infty} P\left[|X_n - X_0| > \epsilon\right] = 0, \ \forall \epsilon > 0$$

This is denoted as $X_n \xrightarrow{P} X_0$

- By the definition of limit, the above means

$$\forall \delta > 0, \ \exists N < \infty, \ s.t. \ P\left[|X_n - X_0| > \epsilon\right] < \delta, \ \forall n > N$$

- We only need marginal distributions of individual X_n to decide whether a sequence converges to a constant in probability

Recap: Weak Law of large numbers

- X_i are iid, $EX_i = \mu$, $\mathsf{Var}(X_i) = \sigma^2$, $S_n = \sum_{i=1}^n X_i$

$$E\left[\frac{S_n}{n}\right] = \mu; \quad \text{and} \quad \mathsf{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}$$

Weak law of large numbers states

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

Recap: almost sure convergence

- A sequence of random variables, X_n, is said to converge **almost surely** or **with probability one** to X if

$$P\left(\{\omega \ : \ X_n(\omega) \rightarrow X(\omega)\}\right) = 1$$

or equivalently

$$P\left(\{\omega \ : \ X_n(\omega) \nrightarrow X(\omega)\}\right) = 0$$

- Denoted as $X_n \stackrel{a.s.}{\rightarrow} X$ or $X_n \stackrel{w.p.1}{\rightarrow} X$ or $X_n \rightarrow X_0 \ \ (w.p.1)$
- We can also write it as

$$P[X_n \rightarrow X] = 1$$

Recap

- The sequence X_n converges to X almost surely iff

$$P\left(\cap_{N=1}^{\infty} \cup_{k=0}^{\infty} \; [|X_{N+k} - X| \geq \epsilon]\right) = 0, \;\; \forall \epsilon > 0$$

Same as

$$P\left(\cap_{N=1}^{\infty} \cup_{k=N}^{\infty} \; [|X_k - X| \geq \epsilon]\right) = 0, \;\; \forall \epsilon > 0$$

- Equivalently

$$\lim_{n\to\infty} P\left(\cup_{k=n}^{\infty} \; [|X_k - X| \geq \epsilon]\right) = 0, \;\; \forall \epsilon > 0$$

- $X_n \stackrel{P}{\to} X$ iff

$$\lim_{n\to\infty} P\left[|X_n - X| > \epsilon\right] = 0, \; \forall \epsilon > 0$$

$$X_n \stackrel{a.s.}{\to} X \quad \Rightarrow \quad X_n \stackrel{P}{\to} X$$

- Almost sure convergence is a stronger mode of convergence

Recap: lim sup and lim inf

- Let $A_1, A_2, \cdots$ be a sequence of events.
- We define

$$\begin{aligned} \lim\ \sup A_n &\triangleq \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k \\ \lim\ \inf A_n &\triangleq \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k \end{aligned}$$

- If $\lim\ \sup A_n = \lim\ \inf A_n$ then that is $\lim A_n$. Otherwise the sequence does not have a limit
- $\lim\ \sup A_n$ and $\lim\ \inf A_n$ are events
- $\lim\ \inf A_n \ \subset\ \lim\ \sup A_n$

Recap

- $X_n \stackrel{a.s.}{\to} X$ iff

$$P\left(\cap_{N=1}^{\infty} \cup_{k=N}^{\infty} \; [|X_k - X| \geq \epsilon]\right) = 0, \;\; \forall \epsilon > 0$$

- Let $A_k^\epsilon = [|X_k - X| \geq \epsilon]$.
- Hence, $X_n \stackrel{a.s.}{\to} X$ iff

$$P\left(\lim\ \sup\ A_n^\epsilon\right) = 0, \;\; \forall \epsilon > 0$$

Recall: Borel-Cantelli Lemma

- **Borel-Cantelli lemma**: Given sequence of events, $A_1, A_2, \cdots$
  1. If $\sum_{i=1}^{\infty} P(A_i) < \infty$, then, $P(\lim\ \sup A_n) = 0$
  2. If $\sum_{i=1}^{\infty} P(A_i) = \infty$ and A_i are independent, $P(\lim\ \sup A_n) = 1$

- Given a sequence X_n we want to know whether it converges to X
- Let $A_k^\epsilon = [|X_k - X| \geq \epsilon]$
- $X_n \stackrel{a.s.}{\rightarrow} X$ if

$$P\left(\lim\ \sup\ A_n^\epsilon\right) = 0, \quad \forall \epsilon > 0$$

- By Borel-Cantelli lemma

$$\sum_{k=1}^{\infty} P(A_k) < \infty \;\;\Rightarrow\;\; P(\lim\ \sup A_k) = 0 \;\;\Rightarrow\;\; X_k \stackrel{a.s.}{\rightarrow} X$$

If A_k are ind

$$\sum_{k=1}^{\infty} P(A_k) = \infty \;\;\Rightarrow\;\; P(\lim\ \sup A_k) = 1 \;\;\Rightarrow\;\; X_k \stackrel{a.s.}{\nrightarrow} X$$

Strong Law of Large Numbers

- Let X_n be iid, $EX_n = \mu$, $\text{Var}(X_n) = \sigma^2$, $S_n = \sum_{i=1}^n X_i$
- We saw weak law of large numbers:

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

- Strong law of large numbers says:

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu$$

- Let $A_n^\epsilon = \left[\left|\frac{S_n}{n} - \mu\right| > \epsilon\right]$
- As we saw, by Chebyshev inequality

$$P\left[\left|\frac{S_n}{n} - \mu\right| > \epsilon\right] \leq \frac{\sigma^2}{n\epsilon^2}$$

- This shows $P(A_n^\epsilon) \to 0$ and thus we get weak law
- To prove strong law using Borel-Cantelli lemma, we need $\sum P(A_n^\epsilon) < \infty$
- Since $\sum_n \frac{\sigma^2}{n\epsilon^2} = \infty$, the Chebyshev bound is not useful
- We need a bound: $P[|\frac{S_n}{n} - \mu|] \leq c_n$ such that $\sum_n c_n < \infty$.

- Let us assume X_i have finite fourth moment

$$\left(\sum_{i=1}^{n}(X_i-\mu)\right)^4 = \sum_{i=1}^{n}(X_i-\mu)^4+\sum_{i}\sum_{j>i}\ \frac{4!}{2!2!}\ (X_i-\mu)^2(X_j-\mu)^2+T$$

Where T represent a number of terms such that every term in it contains a factor like $(X_i-\mu)$
Note that $E[(X_i-\mu)(X_j-\mu)^3]=0$ etc. because X_i are independent.

- Hence we get

$$E\left[\left(\sum_{i=1}^{n}(X_i-\mu)\right)^4\right] = nE[(X_i-\mu)^4]+3n(n-1)\sigma^4 \leq C'n^2$$

- Now we can get, using Markov inequality

$$
\begin{aligned}
P\left[\left|\frac{S_n}{n} - \mu\right| > \epsilon\right] &= P\left[|S_n - n\mu| > n\epsilon\right] \\
&= P\left[\left|\sum_{i=1}^{n}(X_i - \mu)\right| > n\epsilon\right] \\
&\leq \frac{E\left(\sum_{i=1}^{n}(X_i - \mu)\right)^4}{(n\epsilon)^4} \\
&\leq \frac{C' n^2}{n^4 \epsilon^4} = \frac{C}{n^2}
\end{aligned}
$$

- Since $\sum_n \frac{C}{n^2} < \infty$, we get $\frac{S_n}{n} \stackrel{a.s.}{\to} \mu$

- Strong law of large numbers says

$$\frac{S_n}{n} \stackrel{a.s.}{\to} \mu \quad \text{ where } \; S_n = \sum_{i=1}^{n} X_i, \;\; X_i \text{ iid}, \;\; EX_i = \mu$$

- We proved it assuming finite fourth moment of X_i.
- This is only for illustration
- Strong law holds without any such assumptions on moments
- Strong law of large numbers says that sample mean converges to the expectation with probability one.

Convergence in r^{th} mean

- We say that a sequence X_n converges in r^{th} mean to X if $E[|X_n|^r] < \infty$, $\forall n$, $E[|X|^r] < \infty$ and

$$E\left[|X_n - X|^r\right] \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty$$

- Denoted as $X_n \xrightarrow{r} X$
- Consider our old example of binary random variables

$$P[X_n = 1] = \frac{1}{n} \qquad P[X_n = 0] = 1 - \frac{1}{n}$$

- All moments of X_n are finite and we have

$$E\left[|X_n - 0|^2\right] = \frac{1}{n} \rightarrow 0$$

- Hence $X_n \xrightarrow{2} 0$.
- In this example X_n converges in r^{th} mean for all r

- Suppose $X_n \xrightarrow{r} X$. Then, by Markov inequality

$$P\left[|X_n - X| > \epsilon\right] \leq \frac{E\left[|X_n - X|^r\right]}{\epsilon^r} \rightarrow 0$$

- Hence

$$X_n \xrightarrow{r} X \quad \Rightarrow \quad X_n \xrightarrow{P} X$$

- In general, neither of convergence almost surely and in r^{th} mean imply the other.
- We can generate counter examples for this easily.
- However, if all X_n take values in a bounded interval, then almost sure convergence implies r^{th} mean convergence

- Consider sequence X_n where X_n are independent with

$$P[X_n = 0] = 1 - a_n; \quad P[X_n = c_n] = a_n$$

- Assume $a_n \to 0$ so that $X_n \xrightarrow{P} 0$
- By Borel-Cantelli lemma

$$X_n \xrightarrow{a.s.} 0 \quad \Leftrightarrow \quad \sum_n a_n < \infty$$

- For convergence in r^{th} mean we need

$$E\left[|X_n - 0|^r\right] = (c_n)^r \ a_n \quad \to \quad 0$$

- Take $a_n = \frac{1}{n}$ and $c_n = 1$. Then $X_n \xrightarrow{r} 0$ but the sequence does not converge almost surely.
- Take $a_n = \frac{1}{n^2}$ and $c_n = e^n$. Then $X_n \xrightarrow{a.s.} 0$ but the sequence does not converge in r^{th} mean for any r.

- Let $X_n \xrightarrow{r} X$. Then
    1. $E[|X_n|^r] \to E[|X|^r]$
    2. $X_n \xrightarrow{s} X, \ \forall s < r$
- The proofs are straight-forward but we omit the proofs

Convergence in distribution

- Let F_n be the df of X_n, $n = 1, 2, \cdots$. Let X be a rv with df F.
- Sequence X_n is said to converge to X **in distribution** if

$$F_n(x) \to F(x), \ \ \forall x \ \text{ where } F \text{ is continuous}$$

- We denote this as

$$X_n \xrightarrow{d} X, \quad \text{or} \quad X_n \xrightarrow{L} X, \quad \text{or } F_n \xrightarrow{w} F$$

- This is also known as **convergence in law** or weak convergence
- Note that here we are essentially talking about convergence of distribution functions.
- Convergence in probability implies convergence in distribution
- The converse is not true. (e.g., sequence of iid random variables)

Examples

- $X_1, X_2, \cdots$ be iid; uniform over $(0,\ 1)$
- $N_n = \min(X_1, \cdots, X_n)$, $Y_n = nN_n$. Does Y_n converge in distribution?

$$P[N_n > a] = (P[X_i > a])^n = (1-a)^n, \quad 0 < a < 1$$

$$P[Y_n > y] = P[N_n > y/n] = \left(1 - \frac{y}{n}\right)^n, \quad \text{if } \ n > y$$

- Hence for any y

$$\lim_{n \to \infty} P[Y_n > y] = \lim_{n \to \infty} \left(1 - \frac{y}{n}\right)^n = e^{-y}$$

- The sequence converges in distribution to an exponential rv

Examples

- Let $\{X_n\}$ be *iid* with density $f(x) = e^{-x+\theta}, x > \theta > 0$.
- Let $N_n = \min(X_1, \cdots X_n)$. Does N_n converge in probability?
- Guess for limit: θ

$$P[|N_n - \theta| > \epsilon] = P[N_n > \theta + \epsilon] = (P[X_i > \theta + \epsilon])^n$$

$$P[X_i > \theta + \epsilon] = \int_{\theta+\epsilon}^{\infty} e^{-x+\theta}\, dx = e^{-\epsilon}$$

$$P[N_n > \theta + \epsilon] = \left(e^{-\epsilon}\right)^n \to 0, \text{ as } n \to \infty, \ \forall \epsilon > 0$$

- Hence $N_n \xrightarrow{P} \theta$
- Does it converge almost surely?

Examples

- $EX_n = m_n$ and $\text{Var}(X_n) = \sigma_n^2$, $n = 1, 2, \cdots$
- Want a sufficient condition for $X_n - m_n$ to converge in probability
- Note that $E[X_n - m_n] = 0$, and $\text{Var}(X_n - m_n) = \sigma_n^2$, $\forall n$

$$P[|X_n - m_n| > \epsilon] \leq \frac{\sigma_n^2}{\epsilon^2}$$

- Hence, a sufficient condition is $\sigma_n^2 \rightarrow 0$.
- What is a sufficient condition for convergece almost surely?

- We have seen different modes of convergence
- $X_n \xrightarrow{d} X$ iff

$$F_n(x) \to F(x), \ \ \forall x \ \ \text{where } F \text{ is continuous}$$

- $X_n \xrightarrow{P} X$ iff

$$\lim_{n\to\infty} P\left[|X_n - X| > \epsilon\right] = 0, \ \forall \epsilon > 0$$

- $X_n \xrightarrow{r} X$ iff

$$E\left[|X_n - X|^r\right] \to 0 \qquad \text{as} \ \ n \to \infty$$

- $X_n \xrightarrow{a.s} X$ iff

$$P[X_n \to X] = 1 \ \ \text{or} \ \ P[\limsup |X_n - X| > \epsilon] = 0$$

- We have the following relations among different modes of convergence

$$X_n \xrightarrow{r} X \quad \Rightarrow \quad X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{d} X$$

$$X_n \xrightarrow{a.s.} X \quad \Rightarrow \quad X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{d} X$$

- All the implications are one-way and we have seen counter examples
- In general, almost sure convergence does not imply convergence in r^{th} mean and vice versa

- Strong and weak laws of large numbers are very useful examples of convergence of sequences of random variables.
- Given X_i are iid, $EX_i = \mu$, $\text{Var}(X_i) = \sigma^2$, $S_n = \sum_{i=1}^{n} X_i$
  - Weak law of large numbers: $\frac{S_n}{n} \stackrel{P}{\rightarrow} \mu$
  - strong law of large numbers: $\frac{S_n}{n} \stackrel{a.s.}{\rightarrow} \mu$
- Another useful result is the Central Limit Theorem (CLT)
- CLT is about (normalized) sums of of independent random variables converging to the Gaussian distribution

Central Limit Theorem

- Given X_i are iid, $EX_i = \mu$, $\mathsf{Var}(X_i) = \sigma^2$, $n = 1, 2, \cdots$

$$S_n = \sum_{i=1}^{n} X_i \;\;\Rightarrow\;\; ES_n = n\mu,\; \mathsf{Var}(S_n) = n\sigma^2$$

- Given any rv Y, let $Z = \frac{Y - EY}{\sqrt{\mathsf{Var}(Y)}}$
- Then, $EZ = 0$ and $\mathsf{Var}(Z) = 1$.
- Define $\tilde{S}_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ $E\tilde{S}_n = 0$, $\mathsf{Var}(\tilde{S}_n) = 1$, $\forall n$
- Central Limit Theorem states: $\tilde{S}_n \xrightarrow{d} \mathcal{N}(0, 1)$

$$\lim_{n \to \infty} P[\tilde{S}_n \leq a] = \Phi(a) \triangleq \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\; dx$$

- Take X_i iid, $EX_i = 0$, Var$(X_i) = 1$, $n = 1, 2, \cdots$
- $S_n = \sum_{i=1}^{n} X_i$
- Strong law of large numbers implies

$$\frac{S_n}{n} \stackrel{a.s.}{\rightarrow} 0$$

- Central Limit Theorem implies

$$\frac{S_n}{\sqrt{n}} \stackrel{a.s.}{\rightarrow} \mathcal{N}(0, 1)$$

Central Limit Theorem

- Given X_i are iid, $EX_i = \mu$, $\text{Var}(X_i) = \sigma^2$, $n = 1, 2, \cdots$

$$S_n = \sum_{i=1}^{n} X_i \quad \tilde{S}_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

- Central Limit Theorem states: $\tilde{S}_n \xrightarrow{d} \mathcal{N}(0, 1)$
- We use characteristic functions for proving CLT

Characteristic Function

- Given rv X, its characteristic function, ϕ_X, is defined by

$$\phi_X(u) = E\left[e^{iuX}\right] = \int e^{iux}\, dF_X(x) \quad (i = \sqrt{-1})$$

- Since $|e^{iux}| \leq 1$, ϕ_X exists for all random variables

Properties of characteristic function

$$\phi_X(u) = E\left[e^{iuX}\right] = \int e^{iux}\, dF_X(x) \quad (i = \sqrt{-1})$$

- ϕ is continuous; $|\phi(u)| \leq \phi(0) = 1$; $\phi(-u) = \phi^*(u)$
- If $Y = aX + b$, $\phi_Y(u) = e^{iub}\phi_X(ua)$
- If $E|X|^r < \infty$, ϕ would be differentiable r times and

$$\phi^{(r)}(u) = E[(iX)^r e^{iuX}]$$

- Let $\mu_r = E[X^r]$ and let $\nu_r = E[|X|^r]$
- If ν_r is finite, then

$$\phi_X(u) = \sum_{s=0}^{r-1} \mu_s \, \frac{(iu)^s}{s!} + \rho(u) \, \mu_r \, \frac{(iu)^r}{r!}$$

  where $|\rho(u)| \leq 1$ and $\rho(u) \to 1$ as $u \to 0$
- If all moments exist, then

$$\phi_X(u) = \sum_{s=0}^{\infty} \mu_s \, \frac{(iu)^s}{s!}$$

- We denote by ϕ_F characteristic function of df F
- Let F_n be a sequence of distribution functions
- **Continuity theorem**
  - If $F_n \to F$ then $\phi_{F_n} \to \phi_F$
  - If $\phi_{F_n} \to \psi$ and ψ is continuous at zero, then ψ would be characteristic function of some df, say, F, and $F_n \to F$

Characteristic function example

- Let X be binomial rv

$$\begin{aligned}\phi_X(u) = E\left[e^{iuX}\right] &= \sum_{k=0}^{n} {}^nC_k \; p^k \; (1-p)^{n-k} \; e^{iuk} \\ &= \sum_{k=0}^{n} {}^nC_k \; (pe^{iu})^k \; (1-p)^{n-k} \\ &= \left(pe^{iu} + (1-p)\right)^n\end{aligned}$$

- Let $X \sim \mathcal{N}(0,1)$

$$
\begin{aligned}
\phi_X(u) = E\left[e^{iuX}\right] &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\ e^{iux}\ e^{-\frac{x^2}{2}}\ dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}((x-iu)^2 - i^2u^2)}\ dx \\
&= e^{-\frac{u^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}((x-iu)^2}\ dx \\
&= e^{-\frac{u^2}{2}}
\end{aligned}
$$