Recap: Convergence in Probability

- $X_n \xrightarrow{P} X_0 \Leftrightarrow$

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

- By the definition of limit, the above means

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

Recap: Almost sure convergence

- $X_n \stackrel{a.s.}{\to} X \Leftrightarrow P[X_n \to X] = 1 \Leftrightarrow$

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

Same as

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

Equivalently

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

Equivalently

$$P\left(\lim \ \sup \ [|X_k - X| \geq \epsilon]\right) = 0, \ \ \forall \epsilon > 0$$

Recap: 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, $P(\lim\ \sup A_n) = 1$
- Let $A_k^\epsilon = [|X_k - X| \geq \epsilon]$. By Borel-Cantelli lemma, if $\forall \epsilon > 0$,

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

If X_k are ind

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

Recap: Laws of Large numbers

- X_i iid, $EX_i = \mu$, $n = 1, 2, \cdots$
- $S_n = \sum_{i=1}^{n} X_i$
- Weak law of large numbers:

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

- Strong law of large numbers:

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

Recap: Convregence in r^{th} mean

- $X_n \xrightarrow{r} X \Leftrightarrow r^{th}$ moments are finite and

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

- $X_n \xrightarrow{r} X \quad \Rightarrow$
  1. $X_n \xrightarrow{P} X$
  2. $E[|X_n|^r] \rightarrow E[|X|^r]$
  3. $X_n \xrightarrow{s} X, \; \forall s < r$

Recap: Convergence in Distribution

- $X_n \xrightarrow{d} X$
  $\Leftrightarrow \;\; F_n(x) \rightarrow F(x), \;\; \forall x$ where F is continuous
- Does not necessarily imply convergence of pmf/pdf.
- However confergence of pdf's (or pmf's) to a pdf (or pmf) implies convergence in distribution.
- $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$
- $X_n \xrightarrow{d} k \Rightarrow X_n \xrightarrow{P} k$, where k is a constant

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

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

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

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

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

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 $\sigma_n^2 \to 0$.
- 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

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

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

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

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

- 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

Central Limit Theorem

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

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

- Define $\tilde{S}_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$
- Then, $E\tilde{S}_n = 0$, $\mathsf{Var}(\tilde{S}_n) = 1$, $\forall n$
- Central Limit Theorem states: $\tilde{S}_n \stackrel{d}{\rightarrow} \mathcal{N}(0, 1)$

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

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

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

- Central Limit Theorem implies

$$\frac{S_n}{\sqrt{n}} \stackrel{d}{\to} \mathcal{N}(0, 1)$$

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

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

$$\phi^{(r)}(u)\big|_{u=0} = E[(iX)^r e^{iuX}]\big|_{u=0} = i^r\ E[X^r]$$

- Let $\mu_r = E[X^r]$ and let $\nu_r = E[|X|^r]$
- If all moments exist, then

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

- 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 $u \to 0$

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

Characteristic function example

- Let X be binomial rv

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

- Recall $M_X(t) = (pe^t + (1-p))^n$
- We have $\phi_X(u) = M_X(iu)$

- Let $X \sim \mathcal{N}(0,1)$
- We know, $M_X(t) = e^{\frac{t^2}{2}}$
- Hence we get the characteristic function as

$$\phi_X(u) = e^{-\frac{u^2}{2}}$$

- Given X_i iid, $EX_i = \mu$, $\text{Var}(X_i) = \sigma^2$, $S_n = \sum_{i=1}^n X_i$
- Let $\tilde{S}_n = \frac{S_n - ES_n}{\sqrt{\text{var}(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$
- **(Lindberg-Levy) Central Limit Theorem**

$$\lim_{n\to\infty} P\left[\tilde{S}_n \leq x\right] = \lim_{n\to\infty} P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt, \ \ \forall x$$

Proof:

- Without loss of generality let us assume $\mu = 0$.
- We use characteristic function of $\tilde{S}_n$ for the proof.

- $S_n = \sum_{i=1}^n X_i$ and $\tilde{S}_n = \frac{S_n - ES_n}{\sqrt{\text{var}(S_n)}} = \frac{S_n}{\sigma\sqrt{n}}$
- Let ϕ be the characteristic function of X_i. Then

$$\phi_{S_n}(t) = E\left[e^{it\sum_{i=1}^n X_i}\right] = \prod_{i=1}^n E\left[e^{itX_i}\right] = (\phi(t))^n$$

$$\phi_{\tilde{S}_n}(t) = E\left[e^{it\frac{S_n}{\sigma\sqrt{n}}}\right] = \left(\phi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n$$

- Recall that we can expand ϕ in a Taylor series

$$\phi(t) = \sum_{s=0}^{r-1} \mu_s \, \frac{(it)^s}{s!} + \rho(t) \, \mu_r \, \frac{(it)^r}{r!}, \quad \rho(t) \to 1, \text{ as } t \to 0$$

- Here we assume: $EX_i = 0$ and $EX_i^2 = \sigma^2$

$$\phi(t) = 1 \; + \; 0 \; - \; \frac{1}{2} \, \rho(t) \, \sigma^2 \, t^2$$

$$\begin{aligned}
\phi\left(\frac{t}{\sigma\sqrt{n}}\right) &= 1 - \frac{1}{2} \, \rho\left(\frac{t}{\sigma\sqrt{n}}\right) \, \sigma^2 \, \frac{t^2}{\sigma^2 n} \\
&= 1 - \frac{1}{2} \, \frac{t^2}{n} \; + \; \frac{1}{2} \, \frac{t^2}{n} \, \left(1 - \rho\left(\frac{t}{\sigma\sqrt{n}}\right)\right) \\
&= 1 - \frac{1}{2} \, \frac{t^2}{n} \; + \; o\left(\frac{1}{n}\right)
\end{aligned}$$

- Hence we get

$$
\begin{aligned}
\lim_{n\to\infty} \phi_{\tilde{S}_n}(t) &= \lim_{n\to\infty} \left(\phi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n \\
&= \lim_{n\to\infty} \left(1 - \frac{1}{2}\,\frac{t^2}{n} \;+\; o\left(\frac{1}{n}\right)\right)^n \\
&= e^{-\frac{t^2}{2}}
\end{aligned}
$$

  which is the characteristic function of standard normal

- By Continuity theorem, distribution function of $\tilde{S}_n$ converges to that of standard Normal rv

$$
\lim_{n\to\infty} P\left[\tilde{S}_n \leq x\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt, \;\; \forall x
$$

- What CLT says is that sums of iid random variables, when appropriately normalized, would always approach the Gaussian distribution.
- It allows one to approximate distribution of sums of independent rv's
- Let X_i be iid and $S_n = \sum_{i=1}^{n} X_i$

$$P[S_n \leq x] = P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{x - n\mu}{\sigma\sqrt{n}}\right] \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

- Thus, S_n is well approximated by a normal rv with mean $n\mu$ and variance $n\sigma^2$, if n is large

Example

- Twenty numbers are rounded off to the nearest integer and added. What is the probability that the sum obtained differs from true sum by more than 3.
- A reasonable assumption is round-off errors are independent and uniform over $[-0.5,\ 0.5]$
- Take $Z = \sum_{i=1}^{20} X_i$, $X_i \sim U[-0.5,\ 0.5]$, X_i iid.
- Then Z represents the error in the sum.

- $Z = \sum_{i=1}^{20} X_i$, $X_i \sim U[-0.5,\ 0.5]$, X_i iid
- $EX_i = 0$ and $\text{Var}(X_i) = \frac{1}{12}$.
- Hence, $EZ = 0$ and $\text{Var}(Z) = \frac{20}{12} = \frac{5}{3}$

$$
\begin{aligned}
P[|Z| \leq 3] &= P[-3 \leq Z \leq 3] \\
&= P\left[\frac{-3}{\sqrt{\frac{5}{3}}} \leq \frac{Z - EZ}{\sqrt{\text{Var}(Z)}} \leq \frac{3}{\sqrt{\frac{5}{3}}}\right] \\
&\approx \Phi\left(\frac{3}{\sqrt{\frac{5}{3}}}\right) - \Phi\left(\frac{-3}{\sqrt{\frac{5}{3}}}\right) \\
&\approx \Phi(2.3) - \Phi(-2.3) \\
&= 0.9893 - 0.0107 \approx 0.98
\end{aligned}
$$

- Hence probability that the sum differs from true sum by more than 3 is 0.02

- We can approximate binomial rv with Gaussian for large n
- Binomial random variable with parameters n, p is a sum of n independent Bernoulli variables:
  $S_n = \sum_{i=1}^n X_i; \ \ X_i \in \{0, \ 1\}$, $P[X_i = 1] = p$, X_i ind
- Hence we can approximate distribution of S_n by

$$
\begin{aligned}
P[S_n \leq x] &= P\left[\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{x - np}{\sqrt{np(1-p)}}\right] \\
&\approx \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)
\end{aligned}
$$

- For large n, binomial rv is like a Gaussian rv with mean np and variance $np(1-p)$
- The approximation is quite good in practice

- S_n be binomial with parameters n, p

$$P[S_n \leq x] \approx \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$

- For example, with $p = 0.95$

$$P[S_{110} \leq 100] \approx \Phi\left(\frac{100 - 110 * 0.95}{\sqrt{110 * 0.05 * 0.95}}\right) \approx \Phi(-1.97) = 0.025$$

- Since S_n is integer-valued, the LHS above is same for all x between two consecutive integers; but RHS changes
- To get a good approximation, to calculate $P[S_n \leq m]$ one uses $P[S_n \leq m + 0.5]$ in the above approximation formula

- CLT allows one to get rate of convergence of law of large numbers
- Let X_i iid, $EX_i = \mu$, $\text{Var}(X_i) = \sigma^2$, $S_n = \sum_{i=1}^{n} X_i$
- By Law of large numbers, $\frac{S_n}{n} \rightarrow \mu$.
- Now, by CLT

$$
\begin{aligned}
P\left[\left|\frac{S_n}{n} - \mu\right| > \epsilon\right] &= P\left[|S_n - n\mu| > n\epsilon\right] \\
&= P\left[\left|\frac{S_n - n\mu}{\sigma\sqrt{n}}\right| > \frac{n\epsilon}{\sigma\sqrt{n}}\right] \\
&\approx 1 - \left(\Phi\left(\frac{n\epsilon}{\sigma\sqrt{n}}\right) - \Phi\left(-\frac{n\epsilon}{\sigma\sqrt{n}}\right)\right) \\
&= 2\left(1 - \Phi\left(\frac{n\epsilon}{\sigma\sqrt{n}}\right)\right)
\end{aligned}
$$

(because $\Phi(-x) = (1 - \Phi(x))$)

Example: Opinion Polls

- let p denote the fraction of population that prefers product A to product B
- We want to estimate p
- We conduct a sample survey by asking n people
- We want to make a statement such as

$$p = 0.34 \pm 0.07 \textit{ with a confidence of } 95\%$$

- Here, the 0.34 would be the sample mean. The other two numbers can be fixed using CLT

- $X_i \in \{0,\ 1\}$ iid, $EX_i = p$, $S_n = \sum_{i=1}^n X_i$
- Now, by CLT, we have

$$
\begin{aligned}
P\left[\left|\frac{S_n}{n} - p\right| > \epsilon\right] &= P\left[|S_n - np| > n\epsilon\right] \\
&= 2\left(1 - \Phi\left(\frac{n\epsilon}{\sqrt{np(1-p)}}\right)\right)
\end{aligned}
$$

- Suppose we want to satisfy

$$
P\left[\left|\frac{S_n}{n} - p\right| > \epsilon\right] = \delta
$$

- We can calculate any one of ϵ, δ or n given the other two using the earlier equation.
- But we need value of p for it!

- Fortunately, $\sqrt{p(1-p)}$ does not change too much with p
- It attains its maximum value of 0.5 at $p = 0.5$
- It is 0.458 at $p = 0.3$ and is 0.4 at $p = 0.2$
- One normally fixes this variance as 0.5 or 0.45 to calculate the sample size, n.
- There are other ways of handling it

- We have

$$P\left[\left|\frac{S_n}{n} - p\right| > \epsilon\right] = 2\left(1 - \Phi\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)\right)$$

- Suppose $n = 900$ and $\epsilon = 0.025$.
  Let us approximate $\sqrt{p(1-p)} = 0.45$. Then

$$2\left(1 - \Phi\left(\frac{0.025 * 30}{0.45}\right)\right) = 2(1 - \Phi(1.66)) \approx 0.1$$

- If we took $\sqrt{p(1-p)} = 0.5$ we get the value as 0.14
- If we use Chebyshev inequality with variance as 0.5 we get the bound as 0.8
- If we change ϵ to 0.05, then at variance equal to 0.5 the probability becomes about 0.02 while the Chebyshev bound would be about 0.2