Central limit theorem

De Moivre for Bernoulli trials
\[ p = \binom{n}{k} p^k (1-p)^{n-k} \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} \]
more generally?
universal for iid., finite variance

random variable \( X \)
cumulative dist. fn.
\[ F(x) = P(X \leq x) \]

exact characterization: non-dec, right continuous, \[ \to 1 \text{ at } \infty \] and \( 0 \) at \( -\infty \)

\[ \varphi(t) = E(e^{itX}) = \int e^{itx} dF(x) \]
\( \mathbb{R} \) \( \varphi \) is cont. fn.
"characteristic fn.", = Fourier transform
"momentogen fn.", = \( C_f \) valued fn.

same information as moment gen. fn., Laplace transform
but better analytically

\[ E(|X|^k) < \infty \implies \varphi \text{ is } C^k \]
\[ \varphi^{(k)}(t) = \int (ix)^k e^{itx} dF(x) \]

\( \mathcal{F} \text{ ind.} \implies \varphi_{\frac{X}{\sqrt{n}}} = \varphi_{\frac{X}{\sqrt{n}}} \cdot \varphi_{\frac{X}{\sqrt{n}}} \) (convolution = product)
$Y$ uniquely determines $F$ and hence distribution of $X$

**Def.** convergence in distribution:

\[ X_n \rightarrow X \text{ in dist.} \]

\[ F_n(x) \rightarrow F(x) \text{ for all } x \text{ at which } F \text{ cont.} \]

\[ F_n(x) \rightarrow F(x) \text{ for a dense set of } x \text{'s} \]

Note: $X_n \rightarrow X \implies P_n(t) \rightarrow P(t)$ for all $t \in \mathbb{R}$

When can we invert this?

**Def.** Seq. of prob. measures on $\mathbb{R}$ is uniformly tight if $\forall \varepsilon > 0 \exists$ fixed compact set $K \subset \mathbb{R}$ such that all the measures give it prob. $\geq 1 - \varepsilon$.

**Lemma** Every uniformly tight seq. of prob. measures on $\mathbb{R}$ has a subsequence converging in dist.

**Proof:** Let $F_1, F_2, \ldots$ be dist. fn's.

By dia., $\exists$ subseq. converging pointwise on any countable set.

Pick a countable dense subset, get limit $F$ of subseq. on it.

Define by \( F(x) = \inf \{ F(y) : y \geq x \} \).

$F$ is a cumulative distr. fn.
lemma \( F_n, F_k \to \cdot \) unif. tight \( Y_n \to \Phi \) pointwise \( \Rightarrow \) \( F_n \to F \) \( \Phi = \text{char.} F \).

Proof: Convergence of char. fns. implies all subsequences have same limit.

This is enough for CLT (Chebyshev \( \Rightarrow \) unif. tight)

Lévy continuity theorem.
\[ F_n \text{ dist. fns.} \quad Y_n \to \Phi \text{ pointwise for some } \Phi \text{ cont. at } 0 \]

Then \( F_n \) are unif. tight and hence converge in dist.

Proof:
\[
1 - \text{Re} \varphi_n(t) = \int_{\mathbb{R}} (1 - \cos xt) \ dF_n(x)
\]
\[
\frac{1}{\varepsilon^2} \int_0^\varepsilon (1 - \text{Re} \varphi_n(t)) \ dt = \int_{\mathbb{R}} (1 - \frac{\sin \varepsilon x}{\varepsilon x}) \ dF(x)
\]
\[
\geq (1 - \sin 1) \int_{\mathbb{R}} P(1 \leq x_n \leq \frac{1}{\varepsilon})
\]
\[
P(1 \leq x_n \leq \frac{1}{\varepsilon}) \leq \frac{1}{1 - \sin 1} \frac{1}{\varepsilon} \int_0^\varepsilon (1 - \text{Re} \varphi_n(t)) \ dt
\]
Now apply dom. conv., cont. at 0, \( \varphi_n(0) = 1 \)

so unif. tight
Law of Large Number

$X_1, X_2, \ldots$ i.i.d., $E|X| < \infty$

$\implies \frac{X_1 + \cdots + X_n}{n}$ converges to $E X = c$

in dist.

$\varphi(t)$ characteristic function of $X$.

$\varphi(e^{-nt}) = \frac{1}{n} \varphi(e^{-nt})$

$\varphi(t) = (1 + i ct + o(t))$

as $t \to 0$

$(1 + \frac{i ct}{n} + o(t))^n \to e^{-ct}$ characteristic function of $\delta_c$

Need $E|X| < \infty$

Cauchy distribution

density $f(x) = \frac{1}{\pi(1+x^2)}$ $\varphi(t) = e^{-|ct|}$

$\varphi(e^{-nt}) = \varphi(t)$

Averaging $n$ i.i.d. samples is no different from taking a single sample.

Stereographic projection.

Random line

tangent to the Cauchy dist.

Ratio of two independent, identically distributed mean-zero Gaussians = Cauchy.
Central Limit Thm.

\[ EX = 0 \]
\[ \sigma^2 := E X^2 < \infty \]
\[ \implies \frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{\text{normal dist.}} \]

\[ \Phi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2) \]
\[ \implies \Phi(t/\sqrt{n}) \xrightarrow{n} e^{-\sigma^2 t^2/2} \]

Berry-Esseen:

\[ O\left( \frac{E|X|^3}{\sqrt{n}} \right) \] convergence

(Bernoulli trials \( \implies \) got up to constants)

If \( E|X|^3 = \kappa \),
\[ 1 - \frac{\sigma^2 t^2}{2} - i\kappa \frac{t^3}{6} + o(t^3) \]
\[ e^{-\sigma^2 t^2/2} e^{-i\kappa t^3/\sqrt{n}} \]

Finite variance is a strong assumption.

E.g., estimating \( \int_0^1 \frac{dx}{x^a} = \frac{1}{1-a} \)

by Monte Carlo

\( 0 < a < 1 \) finite mean
\( a \geq \frac{1}{2} \implies \) m.f. variance

one event can change running average

heavy tails
Stable Laws.

What can suitably rescaled sums of i.i.d. r.v.'s converge to?

Simple case:

Suppose density satisfies
\[ f(x) \sim \frac{C}{|x|^{1+\alpha}} \quad \text{as} \quad |x| \to \infty \]

\[ 0 < \alpha < 2, \]

Suppose also \( f(x) = f(-x) \).

(Representative hypotheses, for simplicity, but not true in prev. \( \alpha > 0 \)).

\[ \Rightarrow \quad Y(t) = 1 - c' |t|^{\alpha} + o(|t|^{\alpha}) \] \hspace{1cm} (5)

\[ \Rightarrow \quad \frac{X_1 + \ldots + X_n}{\sqrt{n}} \quad \text{converges to constant multiple of r.v.} \]

with \[ \alpha \quad \text{char. r.v.} \quad \text{e}^{-|t|^{\alpha}} \]

What's special about C.L.T.?

1st \( \alpha \) term becomes dominant
as soon as \( \alpha > 2 \)

So have universal behavior, most robust convergence in C.L.T. case.

How do we prove (5)?
Proof of \( \text{(ii)} \):

\[
\int e^{ixt} f(x) \, dx = \int_{|x| \leq b} e^{ixt} f(x) \, dx + \int_{|x| > b} f(x) \, dx + \int_{|x| > b} f(x)(e^{ixt} - 1) \, dx
\]

\(1 + o(t^2)\) from total Mt. of \( f \)

no linear term since \( f(-x) = f(x) \), also \( C^2 \)

\[
2 \int_{x > b} f(x)(\cos(xt) - 1) \, dx
\]

\[
\sim 2C \int_{x > b} \frac{\cos(xt) - 1}{x} \, dx \quad \text{as } b \to \infty
\]

WLOG take \( t > 0 \)

\( u = xt \)

\( du = t \, dx \)

\[
t^{-1 + x} \frac{2C}{\pi} \int_{u > bt} \frac{(\cos u) - 1}{u^{x+1}} \, du
\]

as \( t \to 0 \) get convergent integral.