Consider i.i.d. (independent, identically distributed) random vars \(X_1, X_2, X_3, \ldots\)

\(X_i\) has \(\mu = E[X_i]\) and \(\sigma^2 = \text{Var}[X_i]\)

Consider random variables

\[X_1 + X_2 + \ldots + X_n\]

and

\[\frac{1}{n} \sum_{i=1}^{n} X_i\]

**Law of Large Numbers**

If we observe a random variable \(X\) many times (independently) and take the average, this average will converge to a real number which is \(E(X)\).

Formally, let \(X_1, \ldots, X_n\) be independent, identically distributed random variables with mean \(\mu\).

Define \(A_n = \frac{1}{n} \sum_{i=1}^{n} X_i\) Then for any \(\alpha > 0\) we have

\[P_r(|A_n - \mu| > \alpha) \to 0 \quad \text{as} \quad n \to \infty\]

Proof: Use Chebychev’s inequality.

**the central limit theorem (CLT)**

Consider i.i.d. (independent, identically distributed) random vars \(X_1, X_2, X_3, \ldots\)

\(X_i\) has \(\mu = E[X_i]\) and \(\sigma^2 = \text{Var}[X_i]\)

As \(n \to \infty\),

\[\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma \sqrt{n}} \to N(0, 1)\]

Restated: As \(n \to \infty\),

\[M_n = \frac{1}{n} \sum_{i=1}^{n} X_i \to N\left(\mu, \frac{\sigma^2}{n}\right)\]
CLT applies even to even wacky distributions

CLT in the real world

CLT is the reason many things appear normally distributed

Many quantities = sums of (roughly) independent random vars

**Exam scores:** sums of individual problems

**People's heights:** sum of many genetic & environmental factors

**Measurements:** sums of various small instrument errors

...
Consider i.i.d. (independent, identically distributed) random vars $X_1, X_2, X_3, \ldots$

$X_i$ has $\mu = E[X_i]$ and $\sigma^2 = \text{Var}[X_i]$

As $n \to \infty$,

$$
\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma \sqrt{n}} \to N(0, 1)
$$

Restated: As $n \to \infty$,

$$
M_n = \frac{1}{n} \sum_{i=1}^{n} X_i \to N \left( \mu, \frac{\sigma^2}{n} \right)
$$
Example 1

Number of students who enroll in a class is Poisson (100). Professor will teach course in two sections if more than 120 students enroll. What is the probability of two sections?

When applying approximation, use continuity correction:
Think of $\Pr(X=i) = \Pr(i-0.5 < X < i + 0.5)$

Recall Poisson (100) is sum of 100 independent Poisson(1) random variables, so we can apply CLT.

\[
\Pr(X > 119.5) = \Pr\left(\frac{X - 100}{\sqrt{100}} \geq \frac{119.5 - 100}{\sqrt{100}}\right)
\approx 1 - \Phi(1.95)
\approx 0.0256.
\]

Example 2

If 10 fair die are rolled, find the approximate probability that the sum obtained is between 30 and 40, inclusive, using CLT.

$X_i$ is the value of die # $i$.

$E(X_i) = 7/2$ and $Var(X_i) = 35/12$

\[
E(X) = 35 \quad Var(X) = 350/12
\]

\[
Pr(29.5 \leq X \leq 40.5) = Pr\left(\frac{29.5 - 35}{\sqrt{350/12}} \leq \frac{X - 35}{\sqrt{350/12}} \leq \frac{40.5 - 35}{\sqrt{350/12}}\right)
\approx 2 \cdot \Phi(1.0184) - 1
\approx 0.692.
\]