Central Limit Theorem (CLT)

LLN: If $X_1, \ldots, X_n$ iid (independent and identically distributed)

then $S_n = X_1 + \cdots + X_n$ is $\approx N(\mu, n\sigma^2)$ where $\mu = EX$.

How close is $S_n$ to $\mu$?

As Case: $X_i$ is indep $N(0, \sigma^2)$

$X_1 + X_2 \approx N(0, 2\sigma^2)$

$S_n = X_1 + \cdots + X_n$ is $N(0, n\sigma^2)$

so $\frac{S_n}{\sqrt{n}\sigma}$ is $N(0,1)$, for every $n$. 
e.g. \[ \Pr \left( \frac{S_n}{\sqrt{n}} \in [-1, 1] \right) = \Pr \left( N(0,1) \in [-1, 1] \right) = \Phi(1) - \Phi(-1) \]

this is the same as \[ S_n \in [-\sqrt{n}\sigma, \sqrt{n}\sigma] \]

\( \Phi \) is the normal cdf.

**CLT**

Assume \( X_1, \ldots, X_n \) are iid with \( EX_i = \mu \)
and \( \text{Var}(X_i) = \sigma^2 \). Let \( S_n = X_1 + X_2 + \ldots + X_n \).

Then

\[ \frac{S_n - n\mu}{\sqrt{n} \sigma} \xrightarrow{\text{dist.}} N(0,1) \]

Recall: \( Z_n \xrightarrow{\text{dist.}} N(0,1) \) means that for any \( a, b \)

\[ \Pr \left( Z_n \in [a, b] \right) \rightarrow \Pr \left( N(0,1) \in [a, b] \right) = \Phi(b) - \Phi(a), \]
Qn: Let \( S \) be \( S_{1000} \), sum of 1000 dice.

estimate \( P(S_{r} \in [3300, 3700]) \), i.e., within 200 of \( ES \).

Chebyshev: \( P(|S - ES| \geq 200) \leq \frac{\text{Var}(S)}{200^2} \)

This is the complement of \( S \in (3300, 3700) \).

\[ \text{Var}(S) = 1000 \cdot \text{Var}(X), \quad X = \text{single die}. \]

\[ = 1000 \cdot (35/12) \]

So \( P(|S - ES| > 200) \leq \frac{35000}{12 \cdot 200^2} = 0.07 \ldots \)

\[ P(S \in (3400, 3600)) \geq 1 - \frac{35000}{12 \cdot 100^2} = 0.7 \ldots (1 - 0.3) = 0.7 \ldots \]
CLT for same Qn:

\[ 3300 \leq S \leq 3700 \iff -200 \leq S - 3500 \leq 200 \]

\[ \iff \frac{-200}{\sqrt{1000 \cdot 0.01}} \leq \frac{S - 3500}{\sqrt{1000 \cdot 0.01}} \leq \frac{200}{\sqrt{1000 \cdot 0.01}} \]

\[ \approx N(0,1) \]

\[ P(3300 \leq S \leq 3700) \approx P(a \leq N(0,1) \leq b) \]

\[ = \Phi(b) - \Phi(a) \]

\[ a = b \]

\[ b = \frac{200}{\sqrt{1000 \cdot (35/12)}} = 3.7 \]

for one die: \[ \sigma^2 = 35/12 \]

so this probab. is \( \geq 0.99999 \)

\[ \Phi \left( \frac{15 - 3500}{\sqrt{1000 \cdot 35/12}} \right) - \Phi \left( \frac{-100}{\sqrt{1000 \cdot 35/12}} \right) = \Phi(1.9) - \Phi(-1.9) \]
e.g. without treatment, 30% of patients heal.

Treat 100 patients with experimental drug.

50 got better.

Null hypothesis: the drug has no effect. \( H_0 \)

Find \( P(\bar{Z} \geq 50 \mid H_0) \). If this is small, we rule out \( H_0 \).

If \( H_0 \) holds, then \( \bar{Z} \sim \text{Bin}(100, 0.3) \)

\( \bar{Z} = X_1 + \ldots + X_{100} \) where each \( X_i \) is \( \text{Bern}(0.3) \)

By CLT, \( \frac{\bar{Z} - 30}{\sqrt{100 \cdot 0.21}} \sim N(0, 1) \)

\[
P(\bar{Z} \geq 50) = P\left( \frac{\bar{Z} - 30}{\sqrt{100 \cdot 0.21}} \geq \frac{50 - 30}{\sqrt{100 \cdot 0.21}} \right) \approx 1 - \Phi\left( \frac{50 - 30}{\sqrt{21}} \right)
\]

\[
= 1 - \Phi(4.36)
\]
$1 - \Phi(4.36)$ very small, so $H_0$ very unlikely.