This suggests a way to quantify how close a rv like $\bar{X}_n$ is to a constant like $\mu$:

**Def.** A sequence $\xi_1, \xi_2, \ldots$ of rv is said to converge in probability to a constant $b$ if

for all $\varepsilon > 0$, $\lim_{n \to \infty} P(|\xi_n - b| < \varepsilon) = 1$.

This is denoted $\xi_n \xrightarrow{P} b$.

An immediate consequence of Chebyshev's definition is
(weak) law of large numbers

\[ X_i \sim \text{IID} \] a dist. with mean \( \mu \) and variance \( \sigma^2 < \infty \), \( \overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \)

\[ \overline{X}_n \xrightarrow{p} \mu \sqrt{n} \]

This result has a long history: Gerolamo Cardano (1501–1576) asserted it without proof; Jacob Bernoulli (1654–1705) proved it for \( (X_i, \theta) \sim \text{Bernoulli}(\theta) \) (it took him 20 years to find a correct proof, published posthumously in 1713; Bernoulli thought that this theorem proved the existence of God); Siméon Denis Poisson named it the law of large numbers in 1837.

**Corollary** If \( \overline{X}_n \xrightarrow{p} b \) and \( g(\cdot) \) is continuous at \( \bar{z} = b \), then \( g(\overline{X}_n) \xrightarrow{p} g(b) \).
Central Limit Theorem (CLT)

Example \( \bar{X}_n \sim N(\mu, \sigma^2/n) \), \( \sigma < \infty \) \((i=1, \ldots, n)\)

we know that \( \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \) has mean \( \mu \), variance \( \sigma^2/n \) and is normally distributed, 

so that \( \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \) for all \( n = 1, 2, \ldots \)

Does something like this work for other choices of \( X_i \)?

\[ \bar{X}_n \sim ? \]

A: Yes! Most famous result in all of probability:

Central Limit Theorem

\[ \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \rightarrow \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \]
Careful statement: \( \Xi_1, \Xi_2, \ldots \) a sequence of rv's; let \( F_n \) be the CDF of \( \Xi_n \) 

\[ \text{if there exists a CDF } F^* \text{ such that } \lim_{n \to \infty} F_n(x) = F^*(x), \forall x \text{ at } \]

which \( F^*(x) \) is continuous, then \( \Xi_n \xrightarrow{d} F^* \) \( (\text{in distribution}) \)

CLT: \( \bar{\Xi} \sim \) any dist. with mean \( \mu \) and variance \( 0 < \sigma^2 < \infty \), \( \bar{\Xi}_n \xrightarrow{d} N(\mu, \frac{\sigma^2}{n}) \) 

\[ \frac{\bar{\Xi}_n - \mu}{\sigma} \xrightarrow{d} N(0, 1) \mathrlap{\quad \text{Re CLT}} \]

also has a long history: it was
first demonstrated for \( \mathcal{Z} \sim \text{Bernoulli}(\theta) \) by the French/British mathematician Abraham de Moivre (1667 - 1754) in 1733; almost forgotten until revived by the French mathematician Pierre-Simon de Laplace (1749 - 1827) in 1812; almost forgotten again until 1901, when the Russian mathematician Aleksandr Lyapunov even gave a more general proof; more general proof provided by J.W. Lindeberg (Finnish mathematician (1876 - 1932)) and independently by Paul Lévy (French mathematician (1886 - 1971)) in the early 1920s. The name due to Hungarian-American mathematician George Pólya in 1920.
Example: Contaminated water supply:

\[ E = \text{geoseric concentration} \]
\[ \Xi = \text{lead concentration} \]
\[ R = \frac{\Xi}{\Xi + 3} \] (proportion of contamination due to lead)

\[ E(R) = E\left(\frac{\Xi}{\Xi + 3}\right) \text{ difficult to calculate.} \]

Simulation: Randomly sample pairs \( (\Xi_i, \Xi_i) \) from the joint PDF of \( (\Xi, \Xi) \), calculate \( R_i = \frac{\Xi_i}{\Xi_i + 3} \) and

\[ \bar{R} = \frac{1}{n} \sum_{i=1}^{n} R_i \] (simulation) estimate of \( E(R) \).
Q: How big does n need to be to achieve an accuracy target? By definition

\[ |R_i| = \left| \frac{E_i}{\sqrt{\text{i.i.d.}} E_i} \right| \leq 1 \]

\[ \text{can show that} \]

as a result \[ V(R_i) \leq \frac{1}{4} \]

\[ \text{CLT} \]

 Says that dist. of \( \bar{R}_n \) will be close to Normal for large n, with mean \( E(R_i) \) and variance \( \frac{V(R_i)}{n} = \frac{1}{4n} \)

Suppose we want \( \bar{R}_n \) to differ from \( E(R_i) \) by no more than some tolerance \( T \) with probability at least \( 1-\alpha \) ...
\[ SD = \frac{1}{2\sqrt{n}} \text{ so } \frac{1}{SD} = 2\sqrt{n} \text{ and } \\
-\frac{T}{SD} \leq 2T\sqrt{n} \]

\[
I^{-1}(\frac{1}{2}) = \frac{[E(R_i) - T] - E(R_i)}{SD} = \frac{-T}{SD} \leq 2T\sqrt{n}
\]

From which:

\[ n \geq \left( \frac{I^{-1}(\frac{1}{2})}{2T} \right)^2 \]

For instance, set \( T = 0.005 \) (\( \frac{1}{2} \) of 1%)

and \( d = 0.02 \) to get

\[ n \geq \left( \frac{-2.326}{2(0.05)} \right)^2 = 54,119 \]

Simulation replications needed.

Case Study: Escalators in the London Underground.