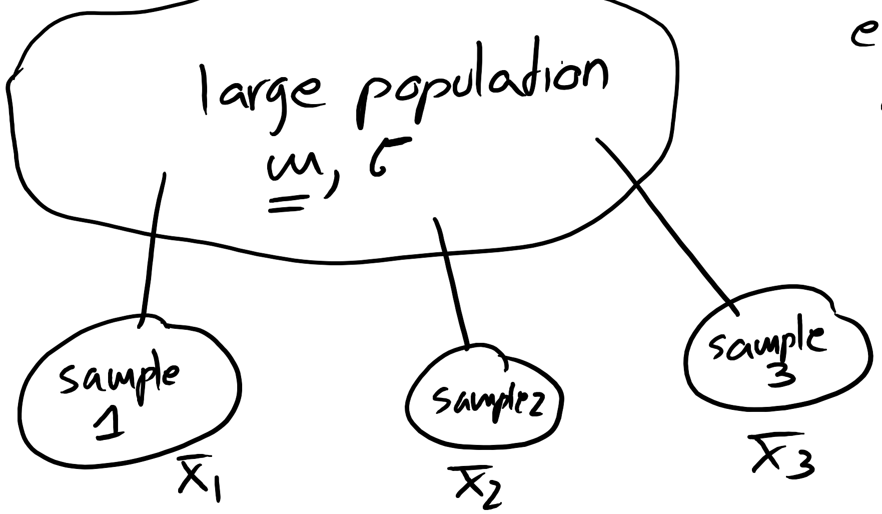
Central limit theorem and the distribution of the sample mean

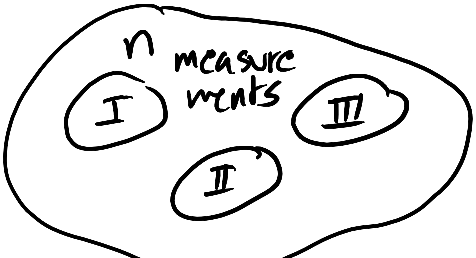


want estimate on μ

Draw samples of size n.

Then $\text{avg}(\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots)$ is an estimate for μ.

But that is the same as combining the measurements into one big sample:



and taking its sample mean $\bar{x}$.

__CLT__ If $n \geq 30$, then $\bar{x}$ has a distribution which is approximately normal.

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

That is $E[\bar{x}] = \mu$ and $Var[\bar{x}] = \frac{\sigma^2}{n}$.

$\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots$ vary but cluster around mean μ.

Using CLT, we can

① conclude something about unknown μ based on $\bar{x}$.

② conclude something about $\bar{x}$ based on μ.

Ex] A company manufactures lightbulbs with mean life 1800 hrs and std dev 200 hrs. (plausible that company has good estimate of these quantities).

(a) Find the probability that a random sample of 100 bulbs will have an avg life of more than 1825 hrs.

$$n \geq 30 \Rightarrow \bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

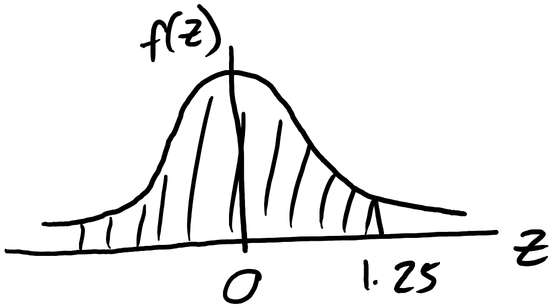
$$P(\bar{x} > 1825) = P\left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} > \frac{1825 - \mu}{\sigma / \sqrt{n}} \right)$$

$$= P\left(z > \frac{1825 - 1800}{(200 / \sqrt{100})} \right)$$

$$= P\left(z > \frac{25}{20} \right) = P(z > 1.25)$$

$$= 1 - P(z \leq 1.25) = 1 - [0.5 + P(0 < z < 1.25)]$$

Note: $z \sim N(0,1)$ (std. normal)



$$= 1 - .8944 = .1056$$

CLT for sample sum

Suppose $X_1, X_2, \ldots, X_n$ are n random variables that are independent and identically distributed.

Then:

$S_n = X_1 + X_2 + \ldots + X_n$ is the sample sum

$$E[S_n] = E[X_1] + \ldots + E[X_n] = n\mu$$

$$Var[S_n] = n\sigma^2 \Rightarrow SD[S_n] = \sqrt{n}\,\sigma$$

CLT states that $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \to N(0,1)$

as $n \to \infty$.