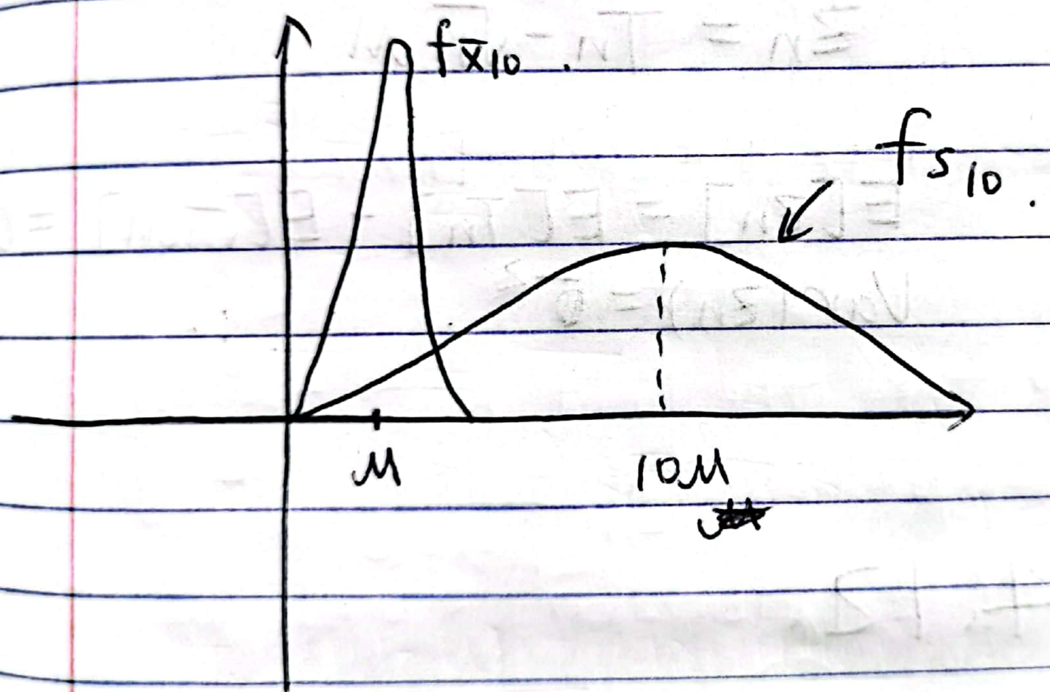
Let us work out the general theory of it.

$X_1, X_2, \cdots X_n$ i.i.d r.v. with mean μ and variance σ^2.

$$S_n = X_1 + X_2 + \cdots + X_n$$

$\Rightarrow E[S_n] = n\mu.$

$\Rightarrow Var(S_n) = n\sigma^2.$



$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$Var(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Consider $T_n = \dfrac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}$.

$$E[T_n] = n \cdot \frac{E[X_i]}{\sqrt{n}} = \sqrt{n} \cdot \mu$$

$$Var(T_n) = \frac{n Var(X_i)}{(\sqrt{n})^2} = \sigma^2.$$

Next time:

$$Z_n = T_n - \sqrt{n}\mu.$$

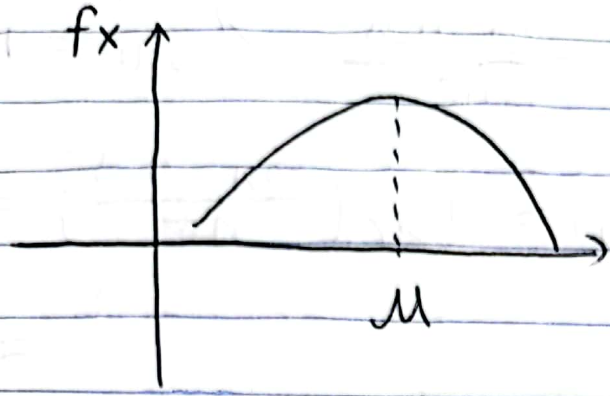
$$E[Z_n] = E[T_n] - E[\sqrt{n}\mu] = 0$$

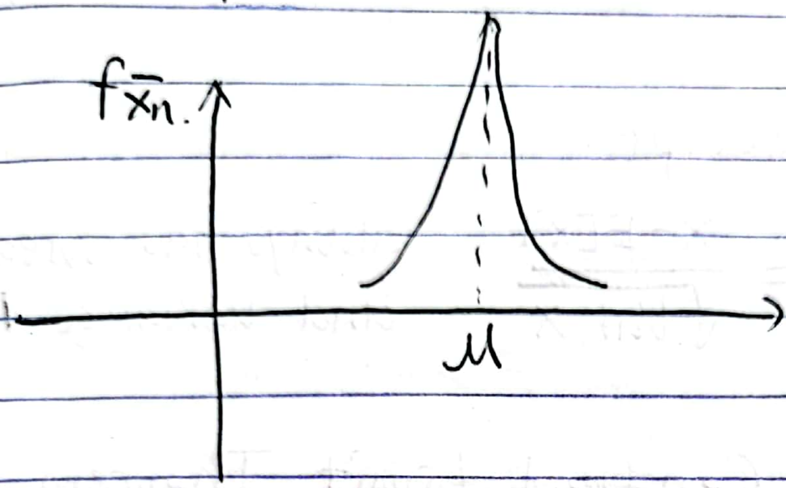
$$Var(Z_n) = \sigma^2.$$

2022.4.12

Last time: Law of large number

$$P(|\bar{X}_n - \mu| > \varepsilon) \longrightarrow 0.$$





$\Rightarrow$ Excellent for μ, but not more information.

Different view points today:

II) The Central Limit Theorem.

(1) Standardizing average.

$$S_n = X_1 + X_2 + \cdots + X_n \qquad E[X_i] = \mu.$$

$$E[S_n] = E[X_1 + X_2 + \cdots + X_n] \qquad Var(X_i) = \sigma^2$$
$$= n\mu.$$

$$Var(S_n) = Var(X_1) + Var(X_2) + \cdots Var(X_n).$$
$$= n\sigma^2.$$

Define $Z_n = \dfrac{S_n - n\mu}{\sqrt{n}\,\sigma} \nearrow$ mean $= 0$, var $= 1$

$$Var(Z_n) = Var\left(\frac{S_n - n\mu}{n\sigma}\right) = \frac{1}{n^2\sigma^2} Var(S_n)$$

$$= \frac{n\sigma^2}{n\sigma^2} = 1$$

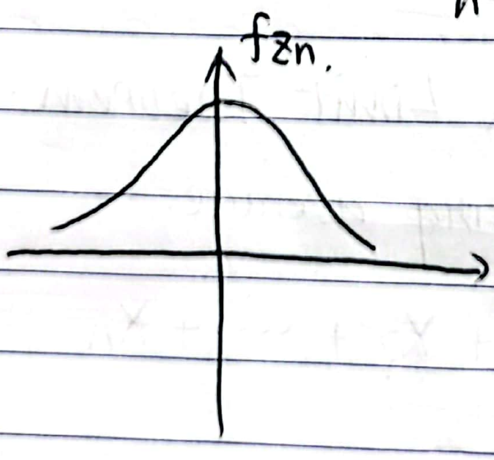
General receipt:

$Z = \dfrac{X - E[X]}{\sqrt{Var\, X}}$ always has mean 0 and variance 1.

(2) The Central Limit Theorem

Now, we take the limit $n \to +\infty$ for Z_n

The CLT says that $Z_n \xrightarrow[n \to +\infty]{} \mathcal{N}(0,1)$.



$$\lim_{n \to +\infty} F_{Z_X}(x) = \Phi(x)$$

In this class, for problems, we say ~~n≥10~~ $n = O(10)$ is a big number.

Other version for Z_n:

$$Z_n = \frac{n\bar{X}_n - n\mu}{\sqrt{n}\,\sigma} = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)$$

Assume we have solved a problem for Z_n, but actually care about $\bar{X}_n$.

~~Example~~:

2) Application of C.L.T.

A) Polling problem from last week.

$X_1, X_2, \cdots\cdots, X_n$ Bernulli r.v. (i.i.d)

$P(|\bar{X}_n - \mu| \geq 0.01) \leq 0.05$.

from last class, we did Chebyshev's inequality:

We found that we would need to poll $\geq 50{,}000$ people.

Here,

$$|\bar{X}_n - \mu| \geq 0.01 \Rightarrow \left|\frac{\sigma \cdot Z_n}{\sqrt{n}} + \mu - \mu\right| \geq 0.01$$

$$\Rightarrow |Z_n| \geq 0.01 \cdot \frac{\sqrt{n}}{\sigma}$$

$$P(|\bar{X}_n - \mu| \geq 0.01) = P\left(|Z_n| \geq 0.01\frac{\sqrt{n}}{\sigma}\right)$$

Last time: $\sigma^2 = \mu(1-\mu)$

$\hookrightarrow \sigma^2 \leq \frac{1}{4}$ conservative

$$P\left(|Z_n| \geq 0.01\frac{\sqrt{n}}{\sigma}\right) \leq P(|Z_n| \geq 0.02\sqrt{n})$$

Want to find n such that:

$$P(|Z_n| \geq 0.02\sqrt{n}) \leq 0.05$$

$$2\times P(Z_n \geq 0.02\sqrt{n}) \leq 0.05.$$

$$P(Z_n \leq 0.02\sqrt{n}) \geq 0.975.$$

Look at tables to find a such that $p(Z_n \leq a) < 0.975$.

$a = 1.96$. from tables.

$$0.02\sqrt{n} = 1.96 \Rightarrow n \approx 9604.$$