14th December, 2022

DSML : CC Maths

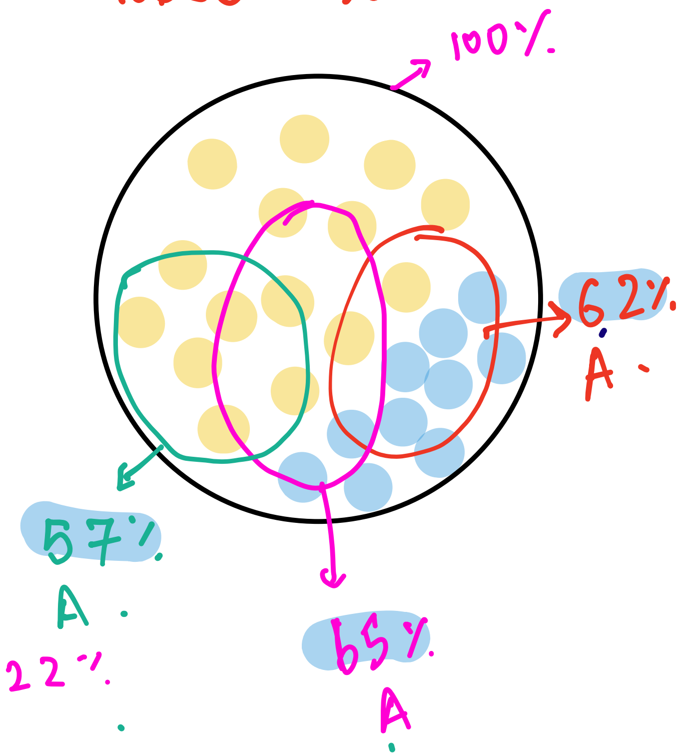
Probability 10 - Confidence Intervals.

Recap: (a) Probability theory.
(b) Bayes' theorem.
(c) Combinatorics.
(d) Descriptive statistics.
(e) Binomial Distribution.
(f) Gaussian Distribution
(g) Poisson Distribution.
(h) Geometric Distribution.
(i) Exponential Distribution.

Class starts @ 9:05 p.m.

Today: (a) Confidence Intervals.
(b) Log-normal Distribution.

Opinion poll



Ground truth: We don't know these.

Candidate A : 60% support.

Candidate B : 40% support.

How to determine the true support for the candidates?

* Not practical to ask everyone.

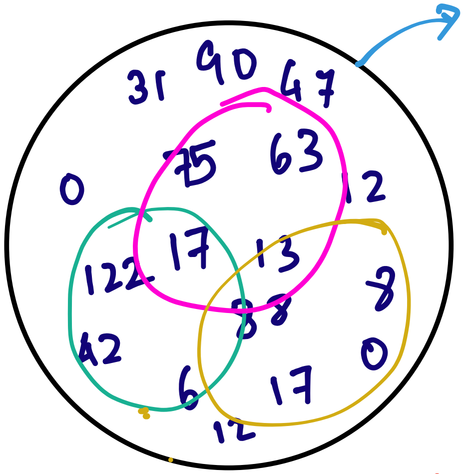
→ Sampling.

* How to make sure that the sample actually reflects the truth? → Choose a large sample.

→ More samples → More accuracy
→ More cost!

Tradeoff.

Sehwag's Runs



Population →

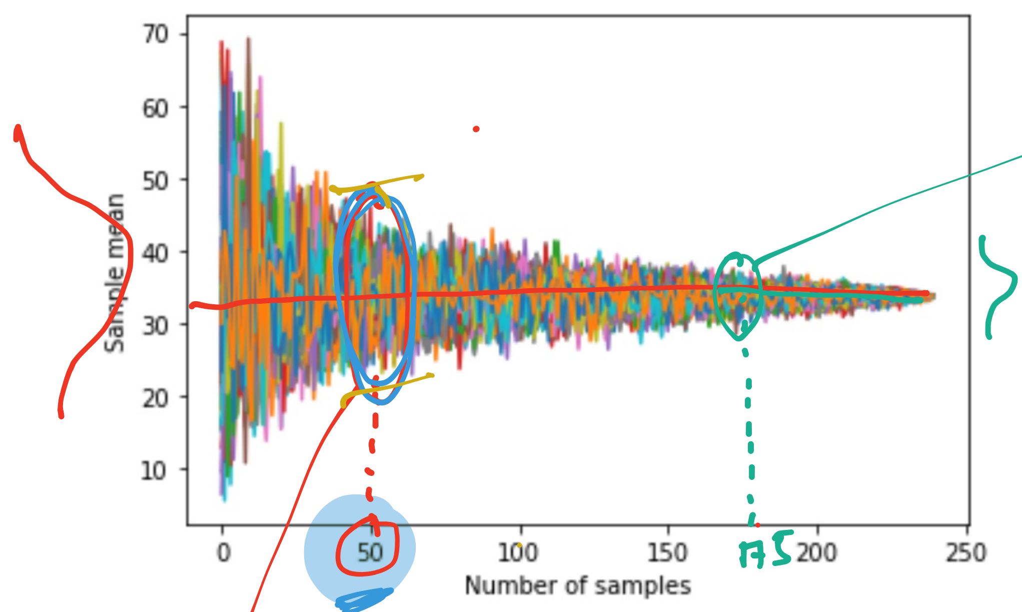
Mean runs : 33.77

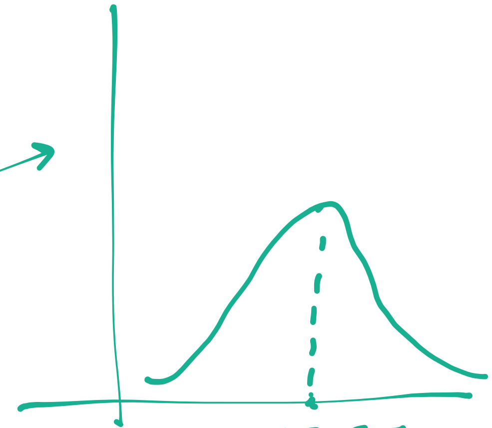
STD runs : 34.81

$\mu_1 = 14.6$	$\sigma_1 = 12.84$
$\mu_2 = 84.6$	$\sigma_2 = 53.36$
$\mu_3 = 45.4$	$\sigma_3 = 34.26$

Sample means.

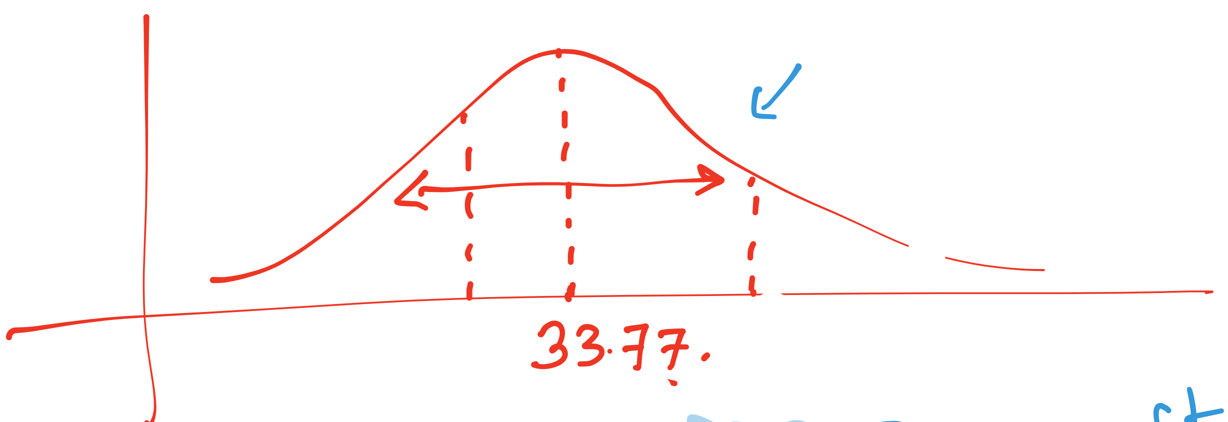
Sample std.





33.77.

Std. dev : $\frac{34.81}{\sqrt{175}}$.



33.77.

Std. dev : $\frac{34.81}{\sqrt{50}}$ } → standard error.

Standard error: If we know the population standard deviation → σ; then the standard error is defined as $\sigma / \sqrt{n}$, where n is the sample size.

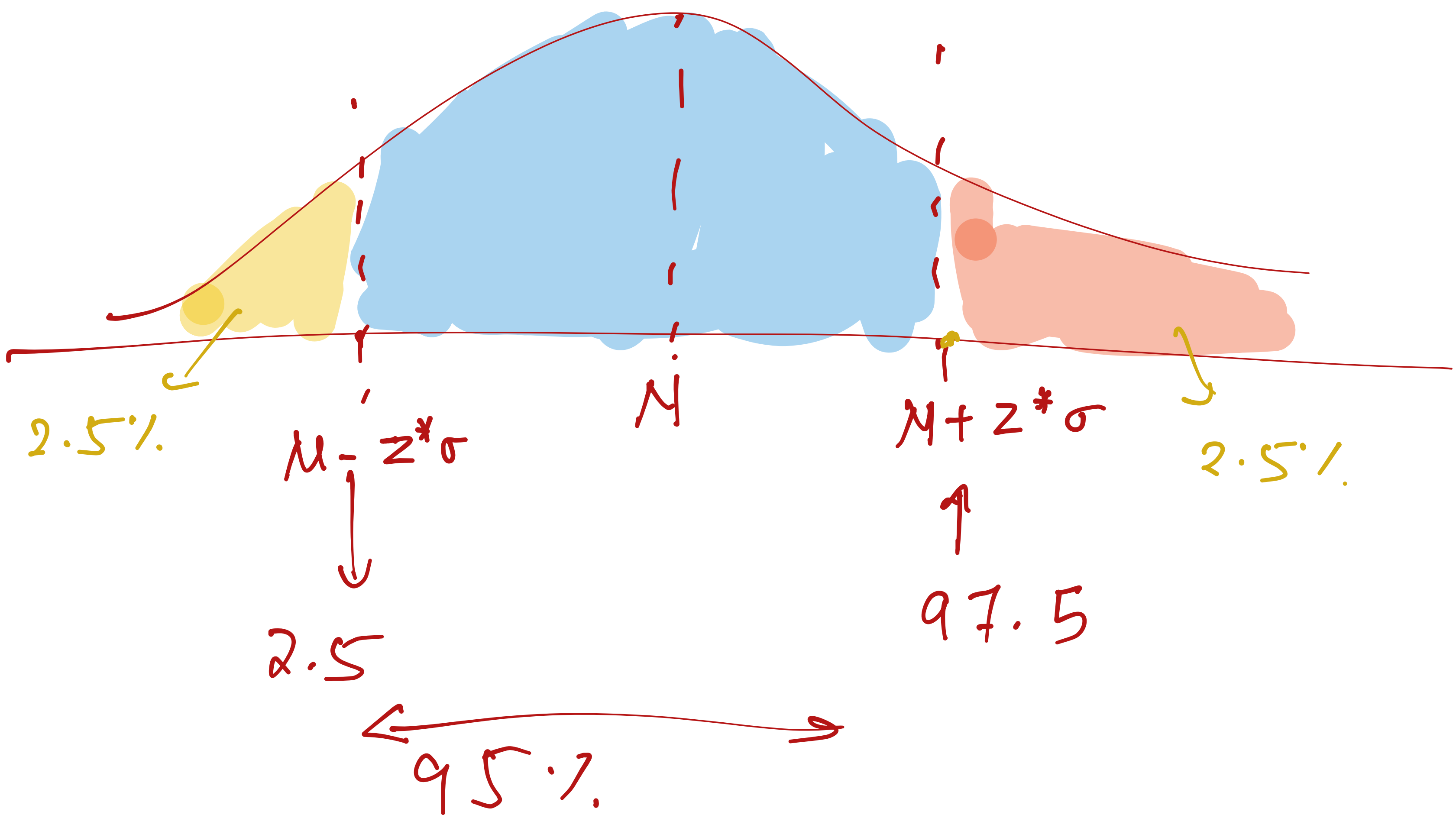
* I want μ of population.
* I collected a sample of size n, and got $\bar{x}$ for that sample.

My confidence interval is defined as:

$$\left[32 - z^* \frac{\sigma}{\sqrt{n}} , 32 + z^* \frac{\sigma}{\sqrt{n}} \right]$$

How to choose z?

Z - z-score.



$Z = -1.96$

$Z = 1.96.$

Confidence Intervals

SDE-2 Salary

You want to know what is the average salary of all SDE-2

Survey 1 Results of a small survey is here

Average = 35 L

Survey 2 Results of another small survey has also come

Average = 35 L

Both surveys have the same mean/average

In which are you more confident? Survey 1

Let us quantify this confidence

Bootstrapping → Confidence Interval.

Survey 1

[35, 36, 33, 37, 34, 35] = 6

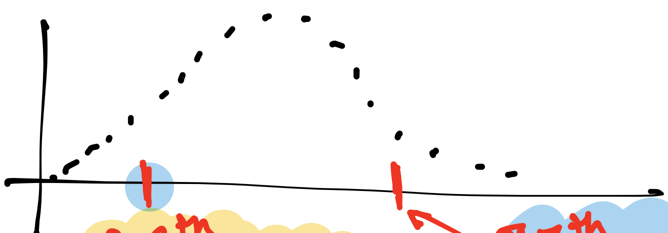
Bootstrapped : [33, 35, 37, 33, 33, 34].

$\bar{x}_1$

$\bar{x}_2$

$\bar{x}_3$

10,000 times.



2.5th percentile

97th percentile → [31 , 39]

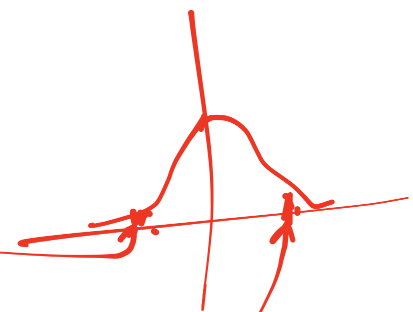
Gaussian: μ : 200

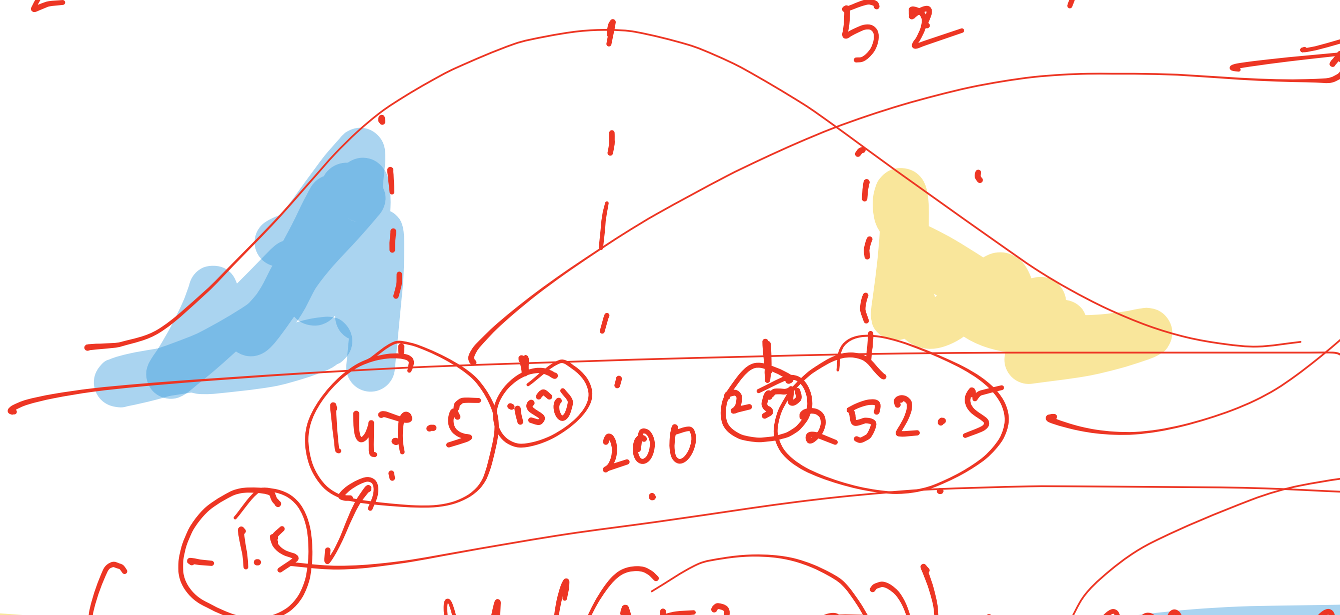
σ : 35.

$$\frac{17.5}{52.5}$$

norm c

$z = 1.5$.





$$\left(1 - \left(\text{norm.cdf}\left(252.5\right)\right)\right) + \text{norm.cdf}\left(147.5\right)$$

1 − no

$$Z = \frac{252.5 - 200}{35} = 1.5$$

$$|X - \mu| > 1.5\sigma.$$

$$\frac{|X - \mu|}{\sigma} > 1.5$$

$$\underbrace{\frac{X - \mu}{\sigma} < -1.5}, \quad \underbrace{\frac{X - \mu}{\sigma} > 1.5.}$$

$\rightarrow X_1 \quad , \quad \rightarrow X_2 .$

$E[X_1] = \mu_1 \qquad E[X_2] = \mu_2 \qquad E[X_1] = \sum x \cdot P(X_1 = x)$

$\sigma_1 \qquad \sigma_2 . \qquad \mu_1 + \mu_2 .$

$(X_1 + X_2) = Y$

$E[X_1 + X_2] = E[Y].$

$\downarrow$

$\sum [x_1 \cdot P[X_1 = x_1] + x_2 \cdot P[X_2 = x_2])$

$\hookrightarrow E[X_1] + E[X_2].$

$$X_1 \quad , \quad X_2$$
$$\downarrow \qquad \downarrow$$
$$\sigma_1 \qquad \sigma_2 .$$

$$\sigma_1^2 = Var(X_1)$$
$$\sigma_1^2 = E[X_1^2] - (E[X_1])^2$$

$$\sigma_2^2 = E[X_2^2] - (E[X_2])^2$$

$$\sigma(X_1 + X_2)?$$

$$Var(X_1 + X_2) = \underbrace{E[(X_1 + X_2)^2]} - (E[\underset{\downarrow}{X_1 + X_2}])^2$$

$$E[X_1^2 + X_2^2 + 2X_1X_2] \qquad (\mu_1 + \mu_2)^2$$
$$E[X_1^2] + E[X_2^2] + 2E[X_1 \cdot X_2] - (\mu_1^2 + \mu_2^2 + 2\mu_1\mu_2)$$

If X_1 and X_2 are independent, then

$$E[X_1 \cdot X_2] = E[X_1] \cdot E[X_2]$$

$$E[X_1^2] + E[X_2^2] + \underbrace{\cancel{2 \cdot E[X_1 \cdot X_2]}}_{} - (\mu_1^2 + \mu_2^2 + \underbrace{\cancel{2\mu_1\mu_2}}_{})$$

applying independence

$$2 \cdot \underbrace{E[X_1]}_{} \cdot \underbrace{E[X_2]}_{}.$$

$$\underbrace{2 \cdot \mu_1 \cdot \mu_2}_{}$$

$$\underbrace{\left(E[X_1^2] - \mu_1^2\right)}_{\sigma_1^2} + \underbrace{\left(E[X_2^2] - \mu_2^2\right)}_{\sigma_2^2}.$$

$Var(X_1 + X_2) = \underline{\sigma_1^2 + \sigma_2^2}. \rightarrow$

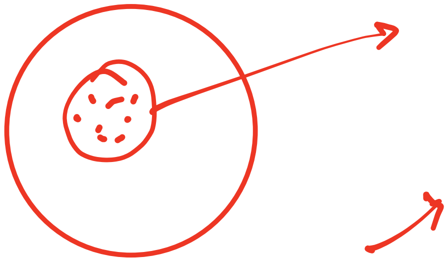
$var(X_1) = \sigma_1^2$

Then, $Var(c \cdot X_1) = \underbrace{c^2 \cdot \sigma_1^2}.$

$Var(X_1 + X_2 + \cdots + \underline{X_n})$

$= Var(X_1) + var(X_2) + Var(X_3) \ldots$

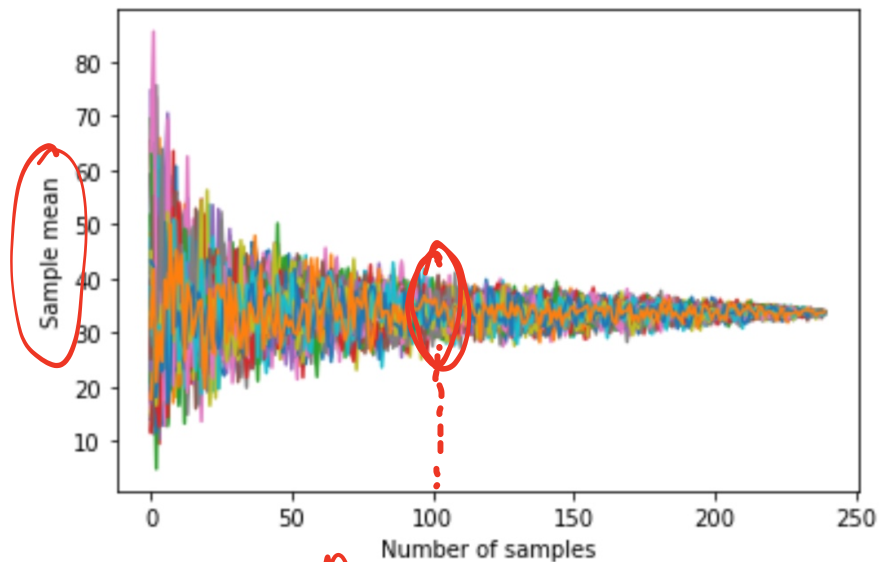
$= \sigma + \sigma + \sigma + \sigma \ldots$ n times



$n \rightarrow (x_1, x_2, x_3, \dots, x_n)$

Let std of all be σ.

$$\bar{X} = \frac{\sum x_i}{n}.$$



$$Var(\bar{X}) = Var\left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i\right)$$

$\frac{1}{n} \rightarrow c$

$$= \frac{1}{n^2} \cdot Var\left(\sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n^{\cancel{2}}} \times \cancel{n}\,\sigma^2 = \frac{\sigma^2}{n}$$

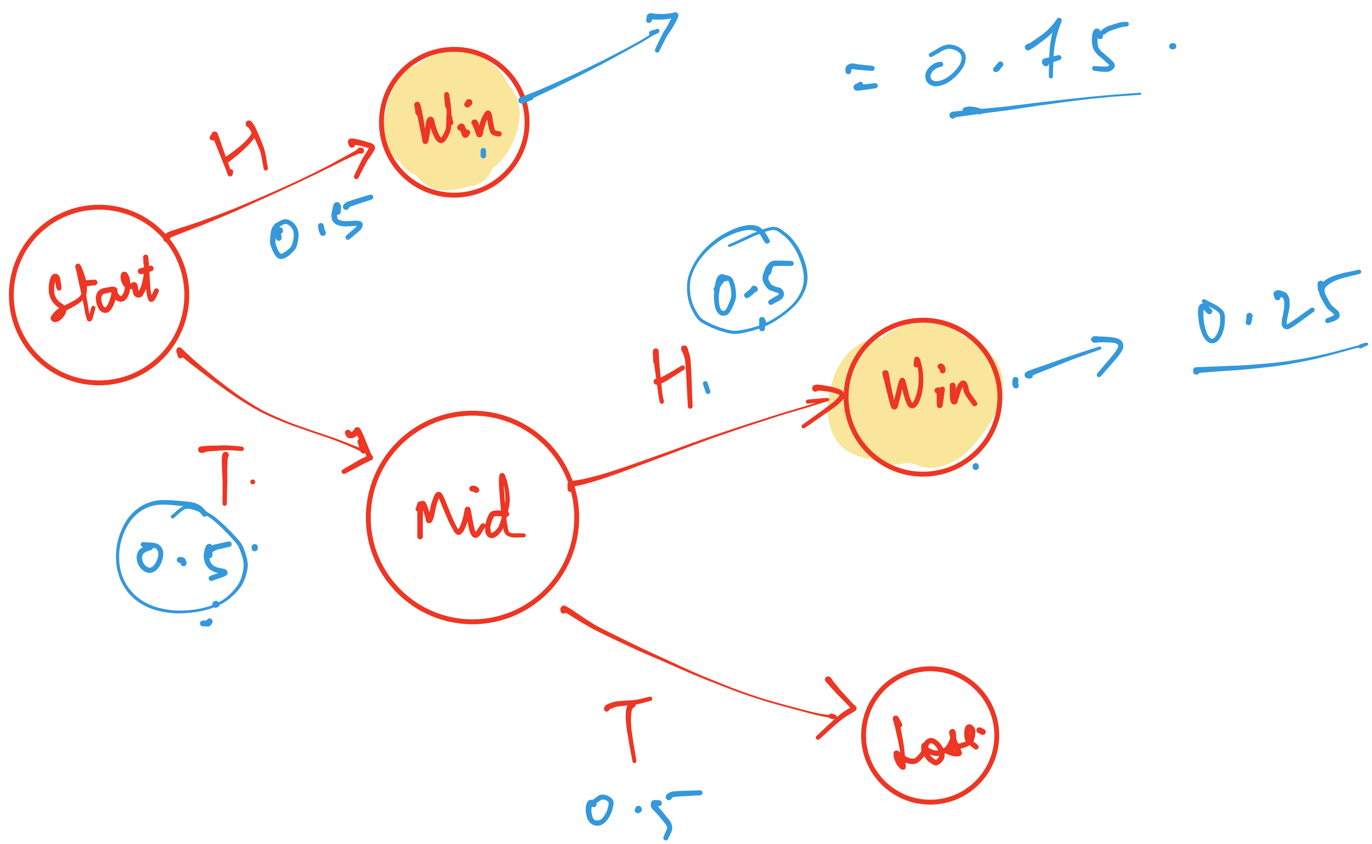
$$\text{std}: \frac{\sigma}{\sqrt{n}}.$$

H → 1$ →

T → 2$ → H.

0.5 + 0.25

= 0.75.



$P(H) = 1/4$.

$\underset{2}{\_\_}\ \underset{2}{\_\_}\ \underset{2}{\_\_} = 8$

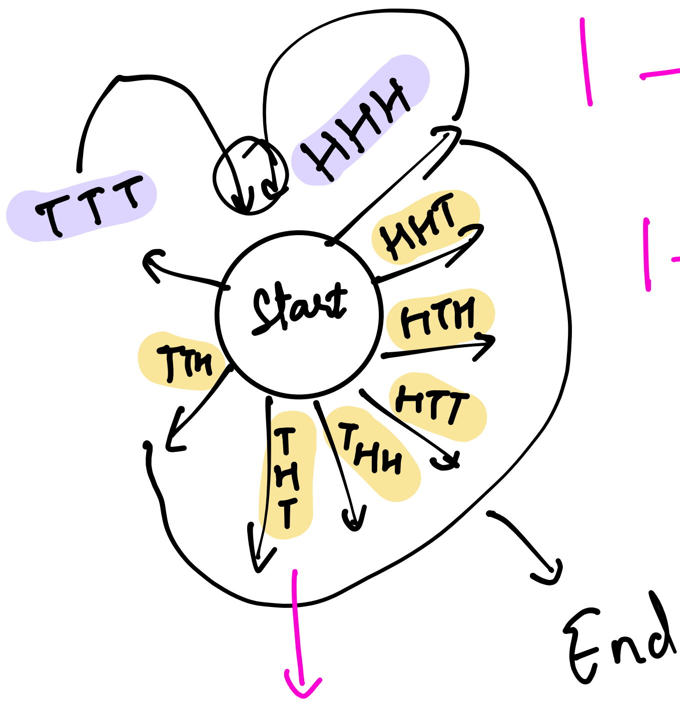
$1 - (P[TTT] + P[HHH])$

$1 - \left(\frac{3^3}{4^3} + \frac{1^3}{4^3}\right)$

$1 - \left(\frac{28}{64}\right)$

$\downarrow$

0.5625.



$P(T) = \frac{3}{4}$.