Topics to be covered:

- inferential statistics - why
- probability
- probability distributions
  - discrete
  - continuous
- statistical inference of
  - point estimate
  - interval estimate

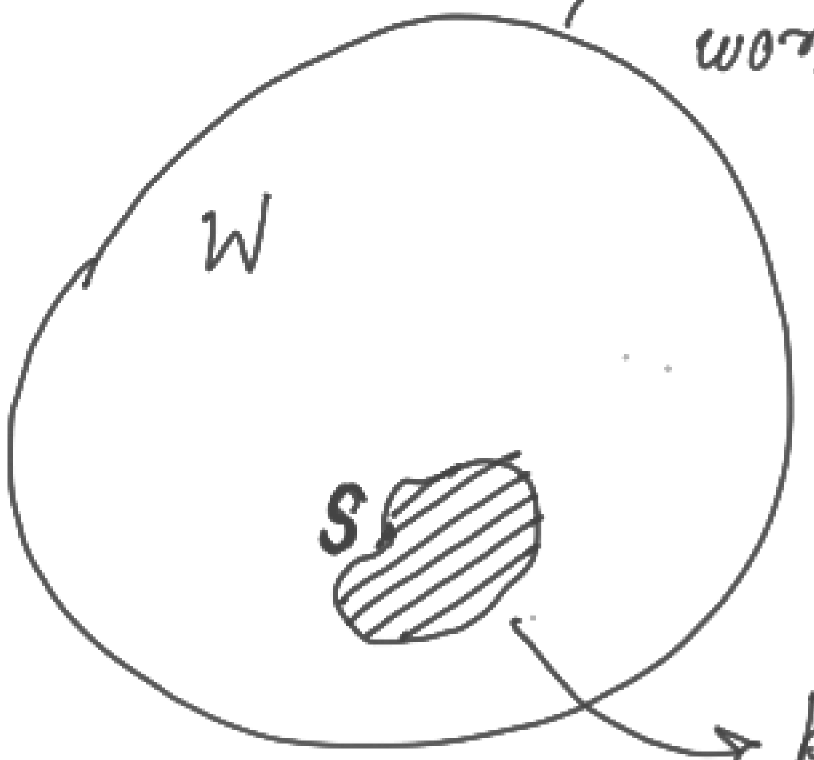
Why inferential statistics?

estimate   property

GOI wants to find avg monthly wage for the working population

(population)

set of all working people in India



people whose monthly wage you know

(sample)

We want to estimate the avg. monthly wage for W by using information about avg. monthly wage in S

By intelligently choosing S; we can estimate properties of W very accurately.

high wage earners

Low wage earners

$\Rightarrow$ Intelligently choosing S would mean doing Stratified Random Sampling instead of Simple Random Sampling

Probability

associated with a random event/ experiment

all outcomes of random experiment are called sample space

an event is a subset of the sample space

RE : A batsman facing a bowler

Sample Space : { 'out', 'dot', 'one', 'two', ... 'seven', 'extras' }

RV : # of runs scored { 0, 0, 1, 2, 7, 1 }

RV : if the batsman got out { 1, 0, 0, 0, . . . , 0 }

Probability (contd...)

Random variable

mapping from the sample space to the set of real numbers

Probability

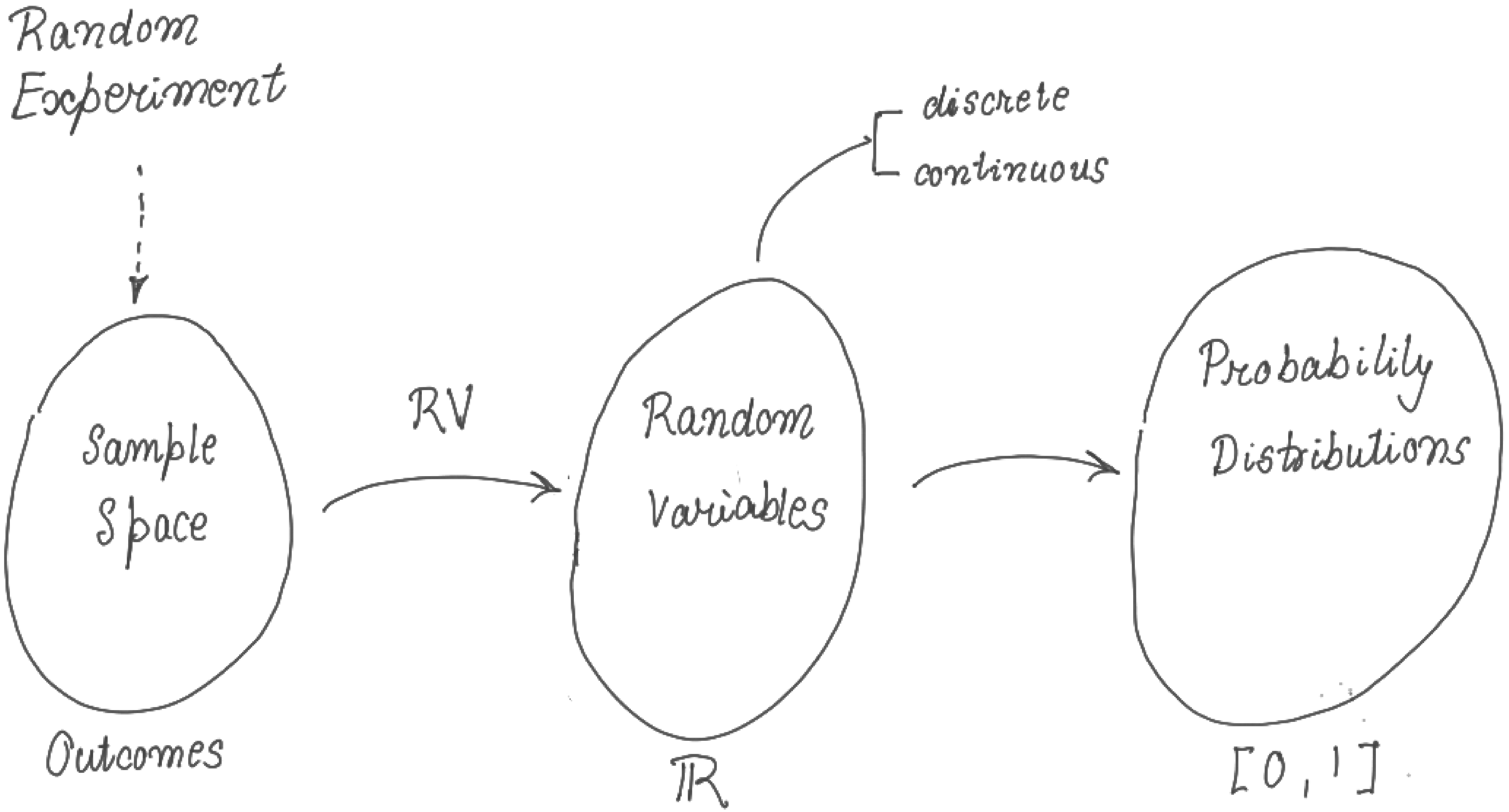
mapping from the set of values of random variables to the set $[0, 1]$

$P(X = x)$ = probability of the event that the random variable X takes the value x.

Axioms of probability

$$1 \geq P(X = x) \geq 0$$

$$\sum_{X} P(X = x) = 1$$


Random Experiment
discrete
continuous
RV
Sample Space
Random Variables
Probability Distributions
Outcomes
$\mathbb{R}$
[0, 1]

Probability Distributions

RE :

SS :

RV: $\{x_1, x_2 \ldots x_k\} \rightarrow$ discrete

probability at a point is 0 by def.^{n} for cont. RV

$[x_1, x_3] \rightarrow$ continuous $\rightarrow$ $P(X = x) = 0$

PDF: $P(X = x_i) = ?$ $\rightarrow D \rightarrow$ dist^n $\rightarrow$ discrete

$P(x_i \leqslant X \leqslant x_j) = ?$ $\rightarrow D \rightarrow$ density $\rightarrow$ continuous

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

$$P(x-h \leqslant X \leqslant x+h) = f(x)$$

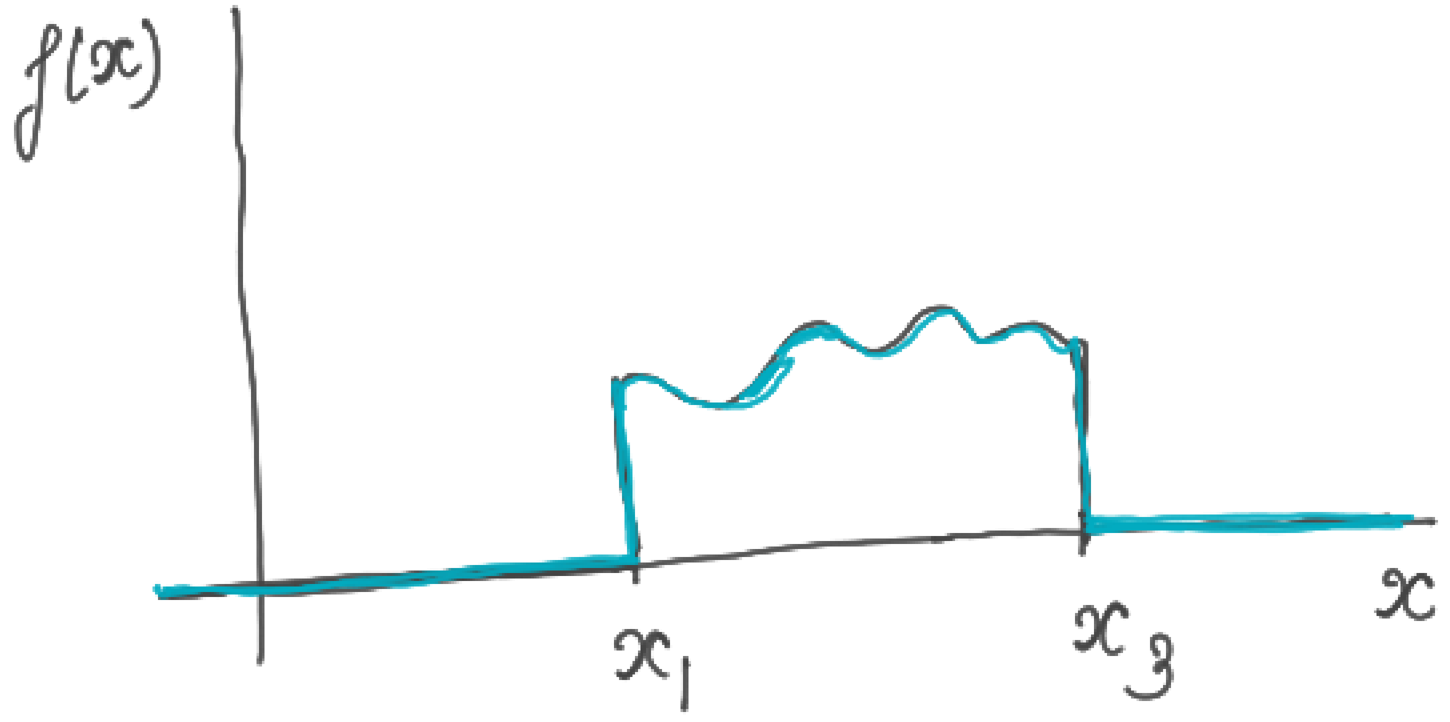
Given the PDF of any probability dist^n; we want to calculate some characteristics:

- central value of the RV $\rightarrow$ Mean $\rightarrow$ $E(X)$
- extent of spread of the RV $\rightarrow$ Variance $\rightarrow$ $var(X)$

$$E(X) = \sum_{X} x\, P(X=x)\, dx = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$var(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \sum_{X} x^2\, P(X=x)\, dx = \int_{-\infty}^{\infty} x^2\, f(x)\, dx.$$



$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{x_1}^{x_3} f(x)\,dx$$

$$\Rightarrow \int_{-\infty}^{x_1} \cancelto{0}{f(x)}\,dx + \int_{x_1}^{x_3} f(x)\,dx$$

$$+ \int_{x_3}^{\infty} \cancelto{0}{f(x)}\,dx$$

$\int_{-\infty}^{\infty}$ is just a lazy way to write $\int$ over all values of x.

Random Exp : Tossing an unfair dice

SS : { ⚀ , ⚁ , ⚂ , ⚃ , ⚄ , ⚅ }

RV : { 1, 2, 3, 4, 5, 6 }

$$\sum_X P(X=x) = 1$$

$$P(X=x) \propto x$$

$$P(X=x) = kx \Rightarrow k(1+2+3+4+5+6) = 1$$

$$\Rightarrow k = 1/21$$

X	unfair $P(X=x)$	fair $P(X=x)$
1	1/21	1/6
2	2/21	1/6
3	3/21	1/6
4	4/21	1/6
5	5/21	1/6
6	6/21	1/6

$$E(X) = \sum_X x P(X=x) = \frac{1^2}{21} + \frac{2^2}{21} + \cdots \frac{6^2}{21}$$

$$= \frac{91}{21} = 4.3\overline{3} \rightarrow \text{unfair.}$$

$$E(X) = \frac{21}{6} = 3.5 \rightarrow \text{fair.}$$

$$= \sum_{x=1}^{6} x \cdot \left(\frac{1}{6}\right) = \frac{1}{6} \sum_{1}^{6} \cancelto{21}{(x)}$$

$$E(X^2) = \sum_{x} x^2 P(X=x)$$

unfair

$$= \sum_{x=1}^{6} \frac{x^3}{21} = \frac{1}{21}\left(\frac{6 \times 7}{2}\right)^2 = 21$$

$$var(X) = E(X^2) - E(X)^2$$

unfair

$$= 21 - 4.3\bar{3}^2$$

$$= 21 - 18.8 = 2.2$$

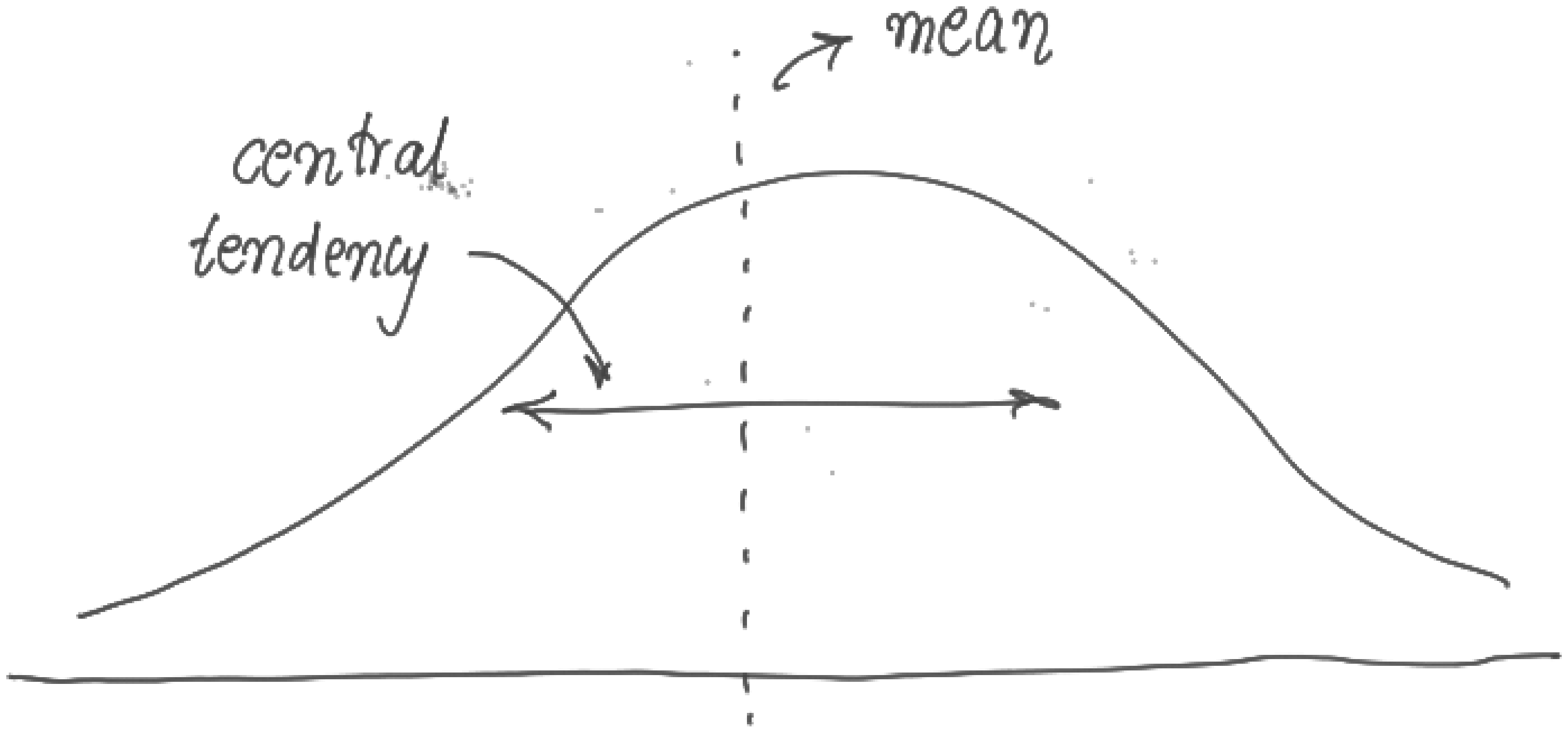
$$E(X^2) = \sum_{x=1}^{6} x^2 \cdot \frac{1}{6} = 15.16$$

fair

$$var(X) = 15.16 - 12.25$$

fair

$$= 2.91$$



Uniform Discrete

$X : \{1, 2, \cdots k\} \rightarrow RV$

PDF : $P(X=x) = \frac{1}{k}$ $\longrightarrow$ fair dice

$E(X) = \frac{(k+1)}{2}$

$var(X) = \frac{k^2-1}{12}$

Bernoulli

$X : \{0, 1\}$ $\rightarrow$ binary outcomes

PDF : $P(X=x) = p^x (1-p)^{1-x}$

$\underline{p} = P(X=1)$

$E(X) = p$ $\quad$ $var(X) = p(1-p)$

Binomial

$X : \{0, 1, \ldots n\}$ $\rightarrow$ x success in n binary outcome exp.

PDF : $P(X=x) = {}^nC_x \, \underline{p}^x (1-p)^{n-x}$

$p = P(X=1)$

$E(X) = np$

$var(X) = np(1-p)$

Geometric

$X : \{0, 1, \ldots\}$ → # of times required to achieve success once in a binary outcome setting

discrete, countable but infinite

$E(X) =$

$var(X) =$

PDF: $P(X=x) = (1-p)^x p$

Poisson

$X : \{0, 1, \ldots\}$ → # of times an event occurs per unit time.

$E(X) = \lambda$

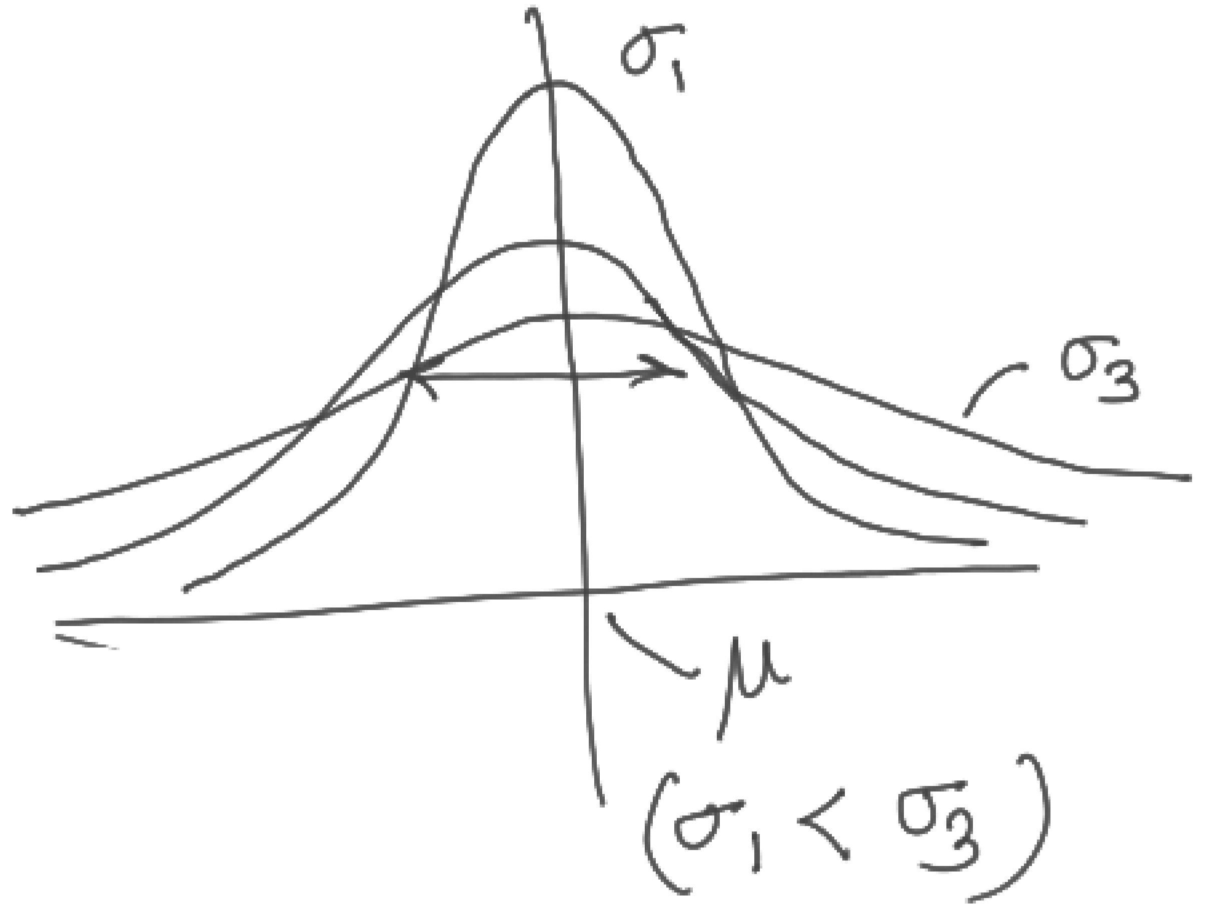
$var(X) = \lambda$

PDF: $P(X=x) = \dfrac{e^{-\lambda} \lambda^x}{x!}$

Normal Dist

$X : (-\infty, \infty)$

PDF : $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



→ natural processes

$E(X) = \mu$ $\quad$ $var(X) = \sigma^2$

↳ Standard Normal Dist $\quad$ $z = \frac{x-\mu}{\sigma}$

$Z : (-\infty, \infty)$

PDF : $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

$E(Z) = 0$

$var(Z) = 1$

$$P(\underbrace{X < \mu + k}) = \int_{-\infty}^{\mu+k} f(x)\, dx$$

difficult ; and calc. redone if μ, σ, k change

SND is a scale invariant sol^n.

$X < \mu + k \Rightarrow Z < k\sigma$

$\int_{-\infty}^{k\sigma} f(z)\, dz$ → looked from table

→ no need to calculate integrals.

Central Limit Theorem.

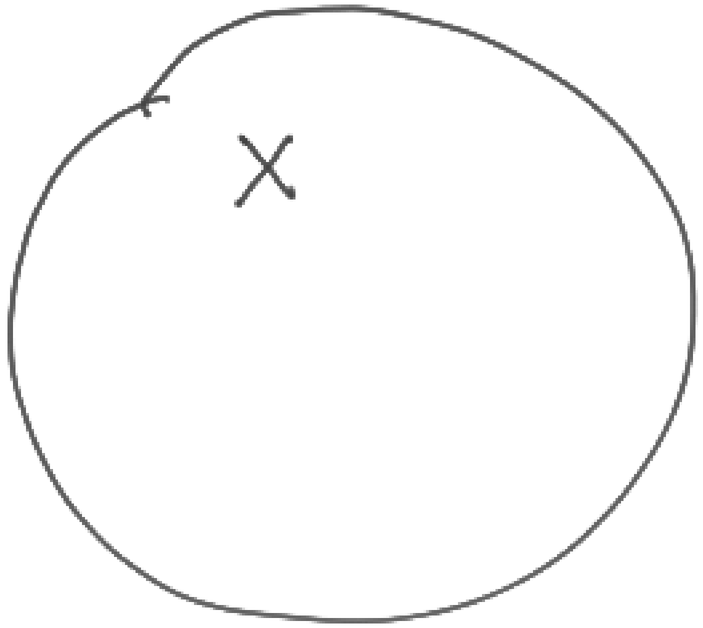
$$\sigma^2 = \frac{\sum \sigma_i^2}{k} \rightarrow t_{k-1}$$

$X \sim D$

$E(X) = \mu$

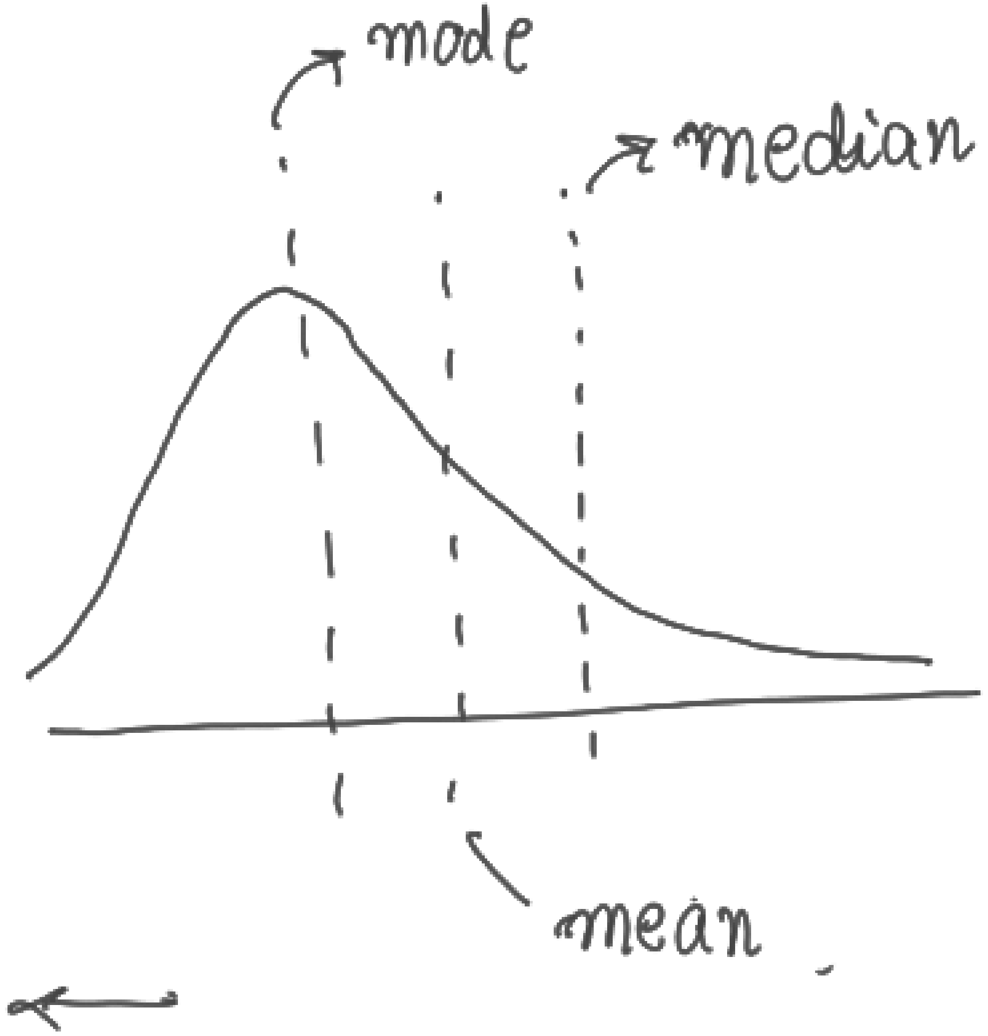
$var(X) = \sigma^2$

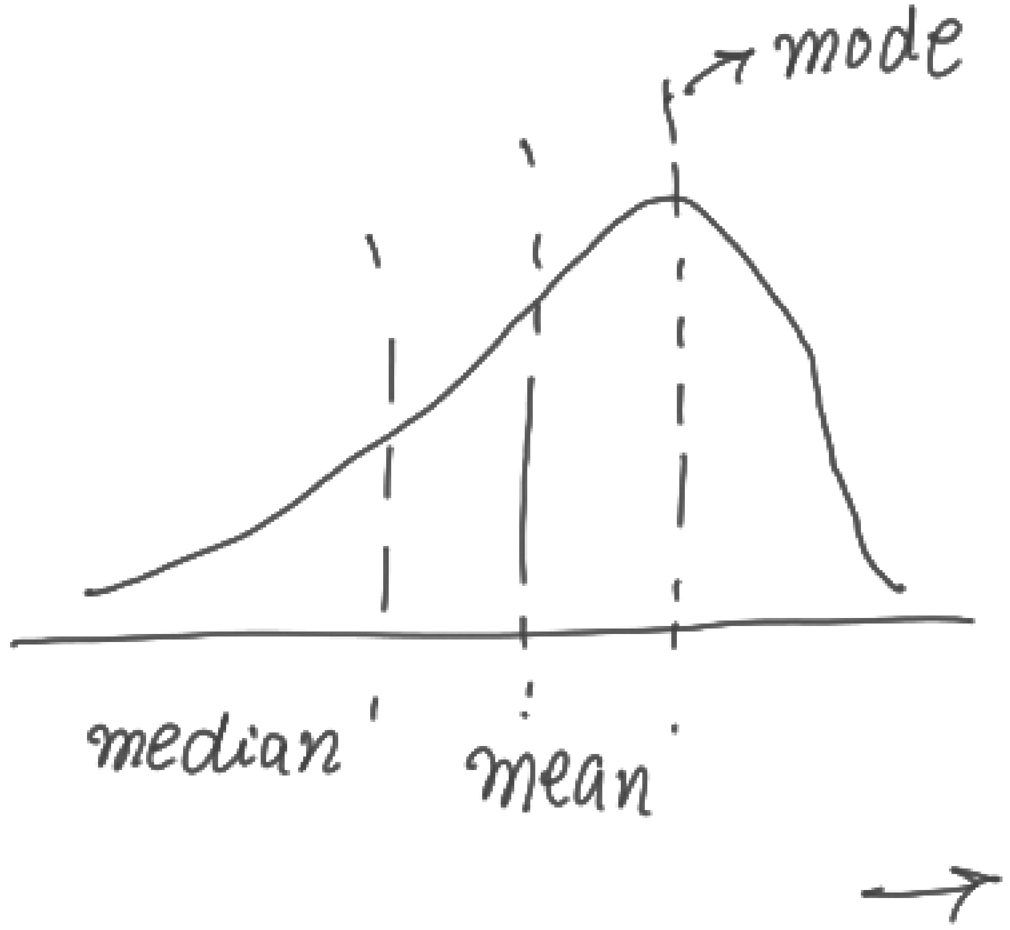
$$\left.\begin{matrix} \sigma_{s_1} \leftarrow s_1 \rightarrow \bar{X}_1 \\ \sigma_{s_2} \leftarrow s_2 \rightarrow \bar{X}_2 \\ \sigma_{s_3} \leftarrow s_3 \rightarrow \bar{X}_3 \\ \vdots \\ \bar{X}_k \end{matrix}\right\} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$



$$\mu \rightarrow \frac{\sum^{k} \bar{X}_i}{k}$$ point estimate

interval estimate $\rightarrow$ $\frac{\sum x_i}{k} - \frac{\sum \sigma_i^2}{k} < \hat{\mu} < \frac{\sum x_i}{k} + \frac{\sum \sigma_i^2}{k}$





skewness $\neq 0$

$\Rightarrow$ non-symmetric PDF