Probability :- is the study of models for (random) experiments when the model is known.

Sample space — random variables — Discrete / Continuous

Statistics :- when the model is not fully known one tries to infer about the "unknown" aspects based on the observed outcomes of the experiment.

Suppose we perform an experiment 'n' times, independent (e.g. Bernoulli trials)

Let $X_1, X_2, \ldots, X_n$ — denote the outcomes of the experiment

- X_i — has the same distribution
  - independent of all other X_j, $j \neq i$

$\{X_i\}_{i \geq 1}$ — i.i.d sequence

i — independent and
i.d — identically distributed.

Q: What can we understand about the "unknown" distribution?

$$\bar{X} = \text{sample mean} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$X_{max} = \text{maximum}\{X_1, \ldots X_n\}$, $X_{min} = \min\{X_1, \ldots X_n\}$

median, Quantiles ...

Summary facts that can be computed

- assumptions that we are making to infer from summary

Some keys to the Question / understanding from Summary :-

Assumption we make is :- Equally probability at each observed outcome.

Empirical distribution : $S = \{X_1, \dots, X_n\}$ ≡ may include repeat observations

is based on discrete distribution on S with probability mass function is given by

$$f(t) = \begin{cases} \frac{1}{n} \# \{i : X_i = t\} & ; t \in S \\ 0 & \text{otherwise} \end{cases}$$

- depend on sample ≡ random
- we don't make any assumption about the underlying distribution.

Inferences based on empirical distribution is called Descriptive statistics. Intuitively, $n \to \infty$, i.e. n gets larger and larger, we will get closer to understanding the true distribution.

let Y be a random variable with p.m.f $f(\cdot)$.

i.e $P(Y=t) = f(t)$ $\quad$ $\text{Range}(Y) = S$.

$$E[Y] = \sum_{t \in S} t \, P(Y=t) = \sum_{t \in S} t \, \underbrace{\frac{\#\{i : X_i = t\}}{n}}$$

$$= \sum_{t \in S} t \, \frac{\sum_{i=1}^{n} 1(X_i = t)}{n}$$

$$= \sum_{i=1}^{n} \frac{1}{n} \sum_{t \in S} t \cdot 1(X_i = t)$$

$$= \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

Theorem (Sample mean) let $X_1, X_2, \ldots X_n$ be an i.i.d. random variables with $E[X_1] = \mu$ and $var[X_1] = \sigma^2$.

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

$$E[\bar{X}] = \mu \qquad var[\bar{X}] = \frac{\sigma^2}{n}$$

Proof:-

$$E[\bar{X}] = E\left[\frac{X_1 + \cdots + X_n}{n}\right]$$

$$= \frac{E[X_1] + \cdots + E[X_n]}{n}$$

$$= \frac{\mu + \cdots + \cdots + \mu}{n} = \frac{n\mu}{n} = \mu$$

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right)$$

Facts about variance

$$= \frac{1}{n^2}\mathrm{Var}(X_1 + \cdots + X_n)$$

X_i - independent

$\mathrm{Var}(X_1+X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)$$

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

Remark :- $E(\bar{X}) = \mu \equiv$ on an average the quantity $\bar{X}$ is accurately describing the unknown μ.

Statistics :- $\bar{X}$ - unbiased Estimator of μ

$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n} \equiv$ as $n \to \infty$ $\mathrm{Var}(\bar{X}) \to 0$,

ie $SD(\bar{X}) \to 0$ as $n \to \infty \Rightarrow$ the "range of $\bar{X}$" gets smaller as $n \to \infty$. Thus with large n, $\bar{X}$ reflects true value, μ, more accurately.

Statistics :- $\bar{X}$ - consistent estimator of μ.

Recall: W - random $E[W] = \mu$, $\mathrm{var}[W] = \sigma^2$

[Tschebychev] $\mathbb{P}(|W-\mu| > k\sigma) \leq \frac{1}{k^2}$

(spelling?)

$\forall k \geq 1$

let $\varepsilon > 0$ be given. Consider the event that

$$\{|\bar{X} - \mu| > \varepsilon\}$$

$E[\bar{X}] = \mu$, $\mathrm{var}[\bar{x}] = \frac{\sigma^2}{n}$, Tschebychev inequality

$$\mathbb{P}(|\bar{X}-\mu| > \varepsilon) = \mathbb{P}\left(|\bar{X}-\mu| > k\frac{\sigma}{\sqrt{n}}\right)$$

$k = \frac{\varepsilon\sqrt{n}}{\sigma}$ and apply Tschebychev to get

$$\leq \frac{1}{k^2} = \frac{\sigma^2}{\varepsilon^2 n}$$

$$\Rightarrow 0 \leq \mathbb{P}(|\bar{X}-\mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n}$$

$\Rightarrow$ For any $\varepsilon > 0$, $\lim_{n \to \infty} \mathbb{P}(|\bar{X}-\mu| > \varepsilon) = 0$

Recall :- $X_1, \dots, X_n$ i.i.d random variables
(– came from n-trials)
of an experiment

- $\overline{X} = \dfrac{X_1 + \dots + X_n}{n}$ – Sample mean

- $f(t) = \dfrac{\#\{i : X_i = t\}}{n}$, Y is r.v. with pmf $f(\cdot)$
then $E[Y] = \overline{X}$ – Empirical distribution.

- $E[\overline{X}] = \mu$ if we assume $E[X_1] = \mu$
$Var[X_1] = \sigma^2$
$Var[\overline{X}] = \dfrac{\sigma^2}{n}$

Theorem : [Weak law of large numbers]
let $n \geq 1$. let $X_1, X_2, \dots, X_n$ be i.i.d random variables. such that $E[X_1] = \mu$ and $var[X_1] = \sigma^2$. Then

$\forall \ \varepsilon > 0$.

$$\mathbb{P}(|\overline{X} - \mu| > \varepsilon) \longrightarrow 0 \quad \text{as } n \to \infty$$

i.e. " $\overline{X} \longrightarrow \mu$ in Probability as $n \to \infty$ "

Typically we are interested in an event A, i.e. $\mathbb{P}(A)$

E.g.:- Toss a coin $-$ $A = \{\text{Head occurs}\}$.

$X_1, \ldots, X_n$, outcome of n-tosses $- \begin{cases} X_i = 1 & \text{if Head occurs} \\ 0 & \text{otherwise} \end{cases}$

$A =$ Head occurs ; $\mathbb{P}(A) = ?$

Define: $\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X} = \frac{\#\text{ of trials in which } A \text{ occurs}}{n}$

$$X_i = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \Rightarrow E[X_i] = 1 \cdot \mathbb{P}(A) + 0 = \mathbb{P}(A)$$

$$E[\hat{p}_n] = E[\overline{X}] = \mathbb{P}(A)$$

$$\mathrm{Var}(\hat{p}_n) = \mathrm{Var}[\overline{X}] = \frac{\mathrm{Var}[X_1]}{n} = \frac{P(A)(1-P(A))}{n}$$

[WLLN] $\forall \varepsilon > 0$

$$\mathbb{P}(|\hat{p}_n - P(A)| > \varepsilon) \longrightarrow 0 \text{ as } n \to \infty$$

- A – event of interest & $\underline{\underline{P(A) = ?}}$

  – Perform an experiment n times $\underset{\text{independent}}{}$ & let

$$X_i = \begin{cases} 1 & \text{if } A \text{ occurs in } i^{th} \text{ trial} \\ 0 & \text{otherwise} \end{cases}$$

$$E[X_i] = P(A) \quad , \quad Var[X_i] = P(A)(1 - P(A))$$

– $$\hat{p}_n = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X} = \frac{\#\text{ of times } A \text{ occurs in } n \text{ trials}}{n}$$

<u>WLLN</u>: $\forall \, \varepsilon > 0$

$$\mathbb{P}\left(|\hat{p}_n - P(A)| > \varepsilon\right) \to 0 \text{ as } n \to \infty$$

relative frequency of A in n trials ·········→ $P(A)$ under the model