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

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

Setup :- Perform an experiment n times. let us be interested in an event A. [unknown model that governs the experiment and provides "$\mathbb{P}(A)$"] ── (**)

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

- Each trial of the experiment is identical and independent of each other. — X_i is identically & distributed independent

Compute : Relative frequency of the occurrence of A

$$= \frac{\# \text{ of times } A \text{ occurs in } n \text{ trials}}{n}$$

$$= \sum_{i=1}^{n} X_i / n \quad \text{———} \quad (*)$$

observations :-

(1) $\cdot\ p = \mathbb{P}(A)$ [True probability of A]

$\cdot$ observed — $X_i = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{o.w.} \end{cases}$

$1 \le i \le n$

relative frequency of A $= \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$

(2) X_i — are all identically distributed & independent $\{X_i\}_{i\ge1}$ i.i.d.

$$E[X_1] = 1 \cdot \mathbb{P}(A) + 0. = \mathbb{P}(A) = p := \mu$$

Given $\{X_i\}_{i\ge1}$ is i.i.d $\Rightarrow E[X_i] = p \quad \forall i \ge 1$

$$E[X_1^2] = p \qquad (\text{Ex})$$

$$\Rightarrow \operatorname{Var}[X_1] = E[X_1^2] - (E[X_1])^2 = p - p^2$$

$$= p(1-p) := \sigma^2$$

Given $\{X_i\}$ is i.i.d $\Rightarrow \operatorname{Var}[X_i] = p(1-p)$.

③ $$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\Rightarrow \quad E[\bar{X}] = \frac{\sum_{i=1}^{n} E[X_i]}{n} = \frac{p \cdot n}{n} = \underset{:=\mu}{p}$$

" on an average $\bar{X}$ is equal to its true value p "

$$\Rightarrow \quad Var[\bar{X}] = Var\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)$$

independence $\longleftarrow$ $$= \frac{\sum_{i=1}^{n} Var[X_i]}{n^2} = \frac{n(p)(1-p)}{n^2}$$

$$= \frac{p(1-p)}{n} := \frac{\sigma^2}{n}$$

" Spread of $\bar{X}$ or range of $\bar{X}$ is getting smaller as n - gets larger " $\longrightarrow$ " $\left(\mu - \frac{\sigma}{\sqrt{n}}, \mu + \frac{\sigma}{\sqrt{n}}\right)$ "

let $\varepsilon > 0$ be given. What is the

$$\mathbb{P}\left(\left| p - \frac{\sum_{i=1}^{n} x_i}{n} \right| > \varepsilon \right) = ?$$

X_i, $i \geq 1$ are all identically distributed & independent

$$\mu = E[X_i] = p$$
$$\sigma^2 = var[X_i] = p(1-p) \quad \forall i \geq 1 \quad ; \quad E[\overline{X}] = \mu, \quad var(\overline{X}) = \frac{\sigma^2}{n}$$

Above is same as:

$$\mathbb{P}\left(|\overline{X} - \mu| > \varepsilon \right) = ?$$

Now above

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

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

[Tschebychev inequality] $\leftarrow$

$$\leq \frac{1}{(\sqrt{n}\,\varepsilon/\sigma)^2} = \frac{\sigma^2}{n\varepsilon^2}$$

$\forall \varepsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}\left(\left| p - \frac{1}{n} \sum_{i=1}^{n} x_i \right| > \varepsilon \right) = 0$$

Theorem: [Weak law of large numbers]

Let $X_1, X_2, \ldots X_n$ be a sequence independent & identically distributed random variables.

Assume X_1 has finite mean μ and finite variance σ^2. Then for any

$\varepsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}\left(|\overline{X} - \mu| > \varepsilon \right) = 0$$

where $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

- Relative frequency ---- Probability ...

-

Theorem [strong law of large numbers]

let $X_1, X_2, \ldots, X_n$ be i.i.d sequence of random variables & $\mu = E[X_1] < \infty$.

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

$$\mathbb{P}\left(\lim_{n \to \infty} \bar{X} = \mu \right) = 1$$

W.L.L.N (weak law of large numbers)

$$\forall \varepsilon > 0 \quad \lim_{n \to \infty} \mathbb{P}\left(|\bar{X} - \mu| > \varepsilon \right) = 0$$

S.L.L.N (strong law of large numbers)

$$\mathbb{P}\left(\lim_{n \to \infty} \bar{X} = \mu \right) = 1$$

Proof of strong law of large numbers

[Mike Kean] — Athreya & Sunder — complete proof

Key steps :- Special [Case :- $X_i \in \{0,1\}$ $E[X_i] = p$ $i \geq 1$]

X_i i.i.d. $i \geq 1$

Step 0 :- $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \in [0,1]$

$\Rightarrow$ $\overline{S} := \limsup_{n\to\infty} \overline{X} \in [0,1]$

$\underline{S} := \liminf_{n\to\infty} \overline{X} \in [0,1]$

$0 \leq \underline{S} \leq \overline{S} \leq 1$

Step 1 :- $\boxed{\text{let } \varepsilon > 0 \text{ be given}}$ For all $k \geq 1$

$$N_k = \inf\left\{ n \geq 1 \;\middle|\; \frac{X_k + \cdots + X_{k+n-1}}{n} \geq \overline{S} - \varepsilon \right\}$$

" N_k – measures how close we are to $\overline{S}$ "

Ex:- N_k is finite with probability 1

Hint: definition of $\bar{S}$

Ex:- $\{N_k\}_{k\geq 1}$ all have the same distribution.

Hint: $\{X_i\}_{i\geq 1}$ are i.i.d.

$\Rightarrow$ $\exists\, m>0$ such that

$$\mathbb{P}(N_k > m) < \varepsilon \quad \forall\, k\geq 1$$

Step 2 :-

For $k\geq 1$,
$$Y_k = \begin{cases} X_k & \text{if } N_k \leq m \\ 1 & \text{if } N_k > m \end{cases}$$

$$N_k^Y = \inf\left\{n\geq 1 \;\middle|\; \frac{Y_k + Y_{k+1} + \dots + Y_{k+n-1}}{n} \geq \bar{S} - \varepsilon\right\}$$

Ex:- (a)- $N_k^Y \leq N_k$ &

(b)- if $N_k > m$ then $N_k^Y = 1$

Hint: $Y_k = 1 \Rightarrow$ we are above $\bar{S}-\varepsilon$ immediately

$\Rightarrow$ (c) $N_k^Y \leq m$ w.p. 1

Step 3 :-

"n-large"

$$\sum_{i=1}^{n} Y_k = \begin{array}{l} Y_1 + \cdots + Y_m + \\ Y_{m+1} \cdots + Y_{2m} + \\ \vdots \\ + \\ \cdots \quad Y_{km+1} + \cdots + Y_n \end{array}$$

$$\geq \begin{array}{l} \bar{S} - \varepsilon + \\ \bar{S} - \varepsilon + \\ \vdots \\ 0 \end{array} = (n-m)(\bar{S}-\varepsilon)$$

$$\Rightarrow \sum_{k=1}^{n} Y_k \geq (n-m)(\bar{S}-\varepsilon) \quad \text{--- } (1)$$

$$\sum_{k=1}^{n} Y_k = \sum_{k=1}^{n} \left(X_k \mathbb{1}_{N_k \leq m} + 1 \cdot \mathbb{1}_{N_k > m} \right) \quad \text{--- } (2)$$

(1) and (2) take Expectation.

$$(n-m)(E(\bar{S}) - \varepsilon) \leq E\left(\sum_{k=1}^{n} Y_k\right)$$

$$\leq \sum_{k=1}^{n} \left[E(X_k \mathbb{1}_{N_k \leq m}) + \mathbb{P}(N_k > m) \right]$$

choice of m $\leq \sum_{k=1}^{n} E[X_k] + n\varepsilon$

X_k i.i.d, $E[X_k] = p$ $\leq np + n\varepsilon$

Ex:- $\Rightarrow E[\bar{S}] \leq p$ ——— (I)

Hint :- . divide above n and let $n \to \infty$

. $\varepsilon > 0$ arbitrary.

step4:

$$\tilde{X}_k = 1 - X_k \quad \forall k \geq 1$$

$\tilde{X}_k \in \{0,1\}$ step1 – step3 for $\tilde{X}_k$ i.i.d

$$\Rightarrow \quad E[\bar{\tilde{S}}] \leq E[\tilde{X}_1]$$

Ex:- $\underline{S} = 1 - \bar{\tilde{S}}$, $E[\tilde{X}_1] = 1 - p$

$$\Rightarrow \quad 1 - E[\underline{S}] \leq 1 - p$$

$$\Rightarrow \quad E[\underline{S}] \geq p \quad \text{— II}$$

step0:

$$\cdot \; 0 \leq \underline{S} \leq \bar{S} \leq 1 \quad -$$

step 3 and step4:

$$E[\bar{S}] \leq p \quad \& \quad E[\underline{S}] \geq p \quad \text{III}$$

Ex: $0 \leq X \leq Y \leq 1$ & $E[X] \geq E[Y] \Rightarrow X = Y$ w.p 1

$$\Rightarrow \quad \bar{S} = \underline{S} \quad \text{w.p. } 1$$

$$\Rightarrow \quad \lim_{n\to\infty} \bar{X} \text{ exists w.p. } 1$$

Step 5: $\underline{\text{WLLN}}$. $\mathbb{P}(|\bar{X} - p| > \varepsilon) \longrightarrow 0$ as $n \to \infty$

$$\Longrightarrow \quad \lim_{n\to\infty} \bar{X} = p \quad \text{w.p. } 1.$$

$\square$

Step 6:- Extension:-

- $\{X_k\}_{k\geq 1}$ with $E[X_1] = \mu < \infty$

by $\quad \bar{X}_k = \begin{cases} \alpha & X_k < \alpha \\ X_k & \alpha \leq X_k \leq \beta \\ \beta & X_k > \beta \end{cases}$

- Step 1 – 5. for $\bar{X}_k$

- limit $\alpha \to -\infty \quad \beta \to \infty$

$\square$