

Recall :-

As n gets larger
Empirical distribution
Captures more information
about True distribution

① X_1, X_2, \dots, X_n i.i.d X

Empirical distribution - discrete distribution
with p.m.f. given by

$$f_n(t) = \frac{1}{n} \# \{i: X_i = t\}$$

② $E[X] = \mu$ and $\text{Var}[X] = \sigma^2$

Y - r.v such that its p.m.f was $f_n(\cdot)$

$$E[Y] = \mu$$

③ ^{strong} Weak-law of large numbers

$$\bar{X} := \frac{X_1 + X_2 + \dots + X_n}{n} \quad \text{then} \quad \bar{X} \xrightarrow{\text{as } n \rightarrow \infty} \mu$$

(WLLN) $\forall \epsilon > 0 \quad \mathbb{P}(|\bar{X} - \mu| > \epsilon) \xrightarrow{\text{as } n \rightarrow \infty} 0$

(SLLN) $A = \{ \bar{X} \rightarrow \mu \text{ as } n \rightarrow \infty \}$

$$\mathbb{P}(A) = 1$$

④ Sample Proportion $p = \mathbb{P}(X \in A)$

$$\hat{p}_n = \frac{\# \{i: X_i \in A\}}{n}$$

as n gets larger $\hat{p} \approx p$
closer

$$z_i = \begin{cases} 1 & \text{if } x_i \in A \\ 0 & \text{if } x_i \notin A \end{cases}$$

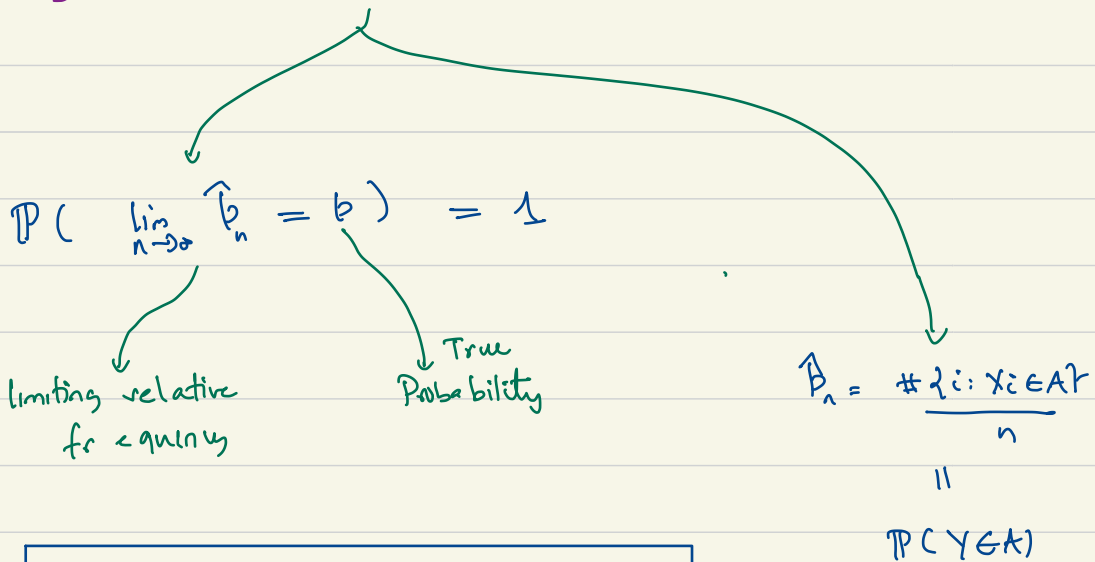
Note:-

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n z_i = \bar{z}$$

- z_i are also independent ; $P(z_i=1) = P(x \in A) = p$
- $z_i \sim \text{Bernoulli}(p)$ $E[z_i] = p$ $\text{Var}[z_i] = 1-p$

[WLLN] $P(|\bar{z} - p| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$

[SLLN] $P(\bar{z} \rightarrow p) = 1$



$E[\bar{x}] = \mu$	- unbiased
$\text{Var}[\bar{x}] = \frac{\sigma^2}{n}$	- Consistent

$\hat{p}_n = \frac{\#\{i: x_i \in A\}}{n}$
 \parallel
 $P(Y \in A)$
 where Y is a r.v. with p.m.f. $f_n(\cdot)$

let X_1, X_2, \dots, X_n i.i.d random variables

$$E[X] = \mu \text{ and } \text{Var}[X] = \sigma^2$$

$$\left[\begin{array}{l} \text{Sample} \\ \text{variance} \end{array} \right] S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n-1}$$

$$E[S^2] = \frac{1}{n-1} \sum_{i=1}^n E[(X_i - \bar{X})^2]$$

$$= \frac{1}{n-1} \sum_{i=1}^n E[X_i^2 - 2X_i \bar{X} + \bar{X}^2]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^n E[X_i^2] - 2 E[\bar{X} (\sum_{i=1}^n X_i)] + n E[\bar{X}^2] \right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^n E[X_i^2] - 2n E[\bar{X}^2] + n E[\bar{X}^2] \right]$$

$$= \frac{1}{n-1} \left[n E[X_i^2] - n E[\bar{X}^2] \right]$$

$$\mu = E[X_i]$$

$$\sigma^2 = \text{Var}[X_i]$$

Similarly

$$E[\bar{X}] = \mu$$

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

$$\Rightarrow \boxed{E[X^2] = \sigma^2 + \mu^2}$$

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

$$E[S^2] = \frac{1}{n-1} [n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)]$$

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

$\Rightarrow S^2$ is an unbiased estimator of σ^2

Ex: $\text{Var}[S^2] \rightarrow 0 \text{ as } n \rightarrow \infty$

Suppose σ^2 is unknown. One can view:

$$(\bar{X} - S, \bar{X} + S)$$

\Rightarrow effective range of X

Simulation :-

Question: Distribution F is given

$$P(X \leq x) = F(x)$$

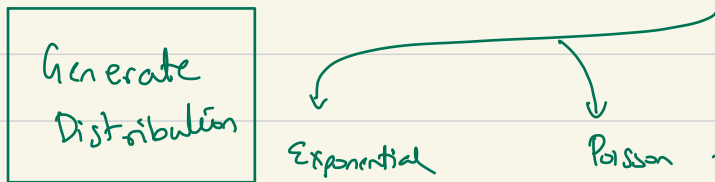
How does one simulate

X_1, X_2, \dots, X_n i.i.d X ?

I_n R -
 $\text{runif}(0,1)$

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x < 1 \\ 1 & x \geq 1 \end{cases} \quad \text{Uniform}(0,1)$$

How does R implement `runif(0,1)` ?



Q: Is there a procedure to go from samples of Uniform(0,1) to any distribution function F ?