Recall :-

① $X_1, X_2, \dots, X_n$ i.i.d X

As n gets larger Emperical distribution captures more information about True distribution

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

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

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

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

$$E[Y] = \mu$$

③ ~~Weak~~ strong - law of large numbers

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

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

(SLLN) $A = \{ \bar{X} \longrightarrow \mu$ as $n \to \infty \}$

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

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

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

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

Note:-

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

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

. z_i are also independent ; $\mathbb{P}(z_i = 1) = \mathbb{P}(x \in A) = p$

$z_i \sim \text{Bernoulli}(p)$ $\quad E[z_i] = p \quad Var[z_i] = 1-p$

[WLLN] $\mathbb{P}(|\bar{z} - p| > \varepsilon) \longrightarrow 0$ as $n \to \infty$

(SLLN) $\mathbb{P}(\bar{z} \longrightarrow p) = 1$

$\mathbb{P}(\lim_{n \to \infty} \hat{p}_n = p) = 1$

limiting relative frequency

True Probability

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

$\|$

$\mathbb{P}(Y \in A)$

where Y is a r.v. with p.m.f. $f_n(\cdot)$

$E[\bar{x}] = \mu$ - unbiased

$Var[\bar{x}] = \frac{\sigma^2}{n}$ - Consistent let $X_1, X_2, \ldots, X_n$ i.i.d random variables

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

[Sample variance] $$S^2 = \frac{(X_1-\bar{X})^2 + (X_2-\bar{X})^2 + \ldots (X_n-\bar{X})^2}{n-1}$$

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

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

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

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

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

$\mu = E[X_1]$ $\Rightarrow$ $\boxed{E[X_1^2] = \sigma^2 + \mu^2}$

$\sigma^2 = Var[X_1]$

Similarly

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

$Var[\bar{X}] = \frac{\sigma^2}{n}$

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

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

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

Ex: $Var[S^2] \longrightarrow 0$ as $n \to \infty$

Suppose σ^2 is unknown. One can view:

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

as effective range of X

Simulation :-

Question: Distribution F is given

$$\mathbb{P}(X \le x) = F(x)$$

How does one simulate

$X_1, X_2, \ldots X_n$ i.i.d X ?

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

$$F(x) = \begin{cases} 0 & x \le 0 \\ x & 0 < x < 1 \\ 1 & x \ge 1 \end{cases}$$

Uniform (0,1)

How does R implement runif(0,1) ?

$\boxed{\text{Generate Distribution}}$

Exponential

Poisson .

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