vi) Box Plot:-

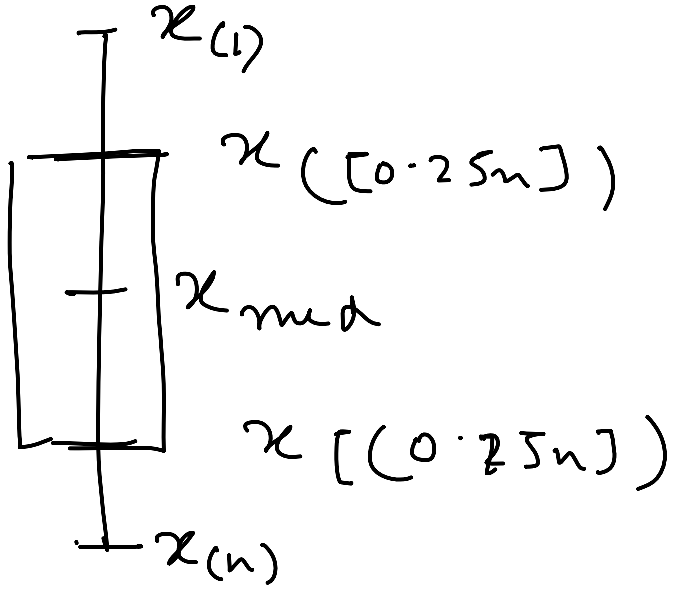
$\mathcal{X} = \{x_1, \dots, x_n\} \to$ data.

Step 1:- Compute $x_{(1)}$ [minimum one] & $x_{(n)}$ [maximum one].

Step 2:- Compute interquartile range [i.e, $x_{([0.25n])}$ & $x_{([0.75n])}$]

Step 3:- Compute median as well. n: Sample size.

Plot



## Section 9 (Estimation) :-

Statistical Inference

Estimation | Testing of hypothesis
(Interval estimation is a part of that).

Estimation

Parametric estimation — included.
(The form of the distⁿ is known but the parameters involved in the distⁿ are unknown. We have to estimate those)

Non-parametric estimation (Not included)
(Don't know anything, have to estimate the underlying distⁿ / density).

**Some basic :-**

$X$ : Random variable.

$p(x) = P[X = x] \longrightarrow$ p.m.f. of $X$ at $x$.

$f_X(x)$ : p.d.f of $X$ at $x$.

$P[a < X < b] = \int_a^b f_X(x)\,dx$.

Parametric estimation:– Suppose, the population is a member of family of p.d.f-s, $\{f(x;\theta) : \theta \in \Theta\}$

Here $\theta$: parameter (unknown) ; $\Theta$: Parameter space.

Example:– 1) Normal family : $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$.

Here $\theta = (\mu, \sigma^2)$ , $\Theta = \mathbb{R} \times \mathbb{R}^+$

2) Poisson family : $\left\{ \dfrac{e^{-\lambda} \lambda^x}{x!} : \lambda > 0 \right\}$.

Here $\theta = \lambda$ & $\Theta = \mathbb{R}^+$.

It is parametric estimation since we want to estimate the unknown parameter $\theta$ or any $f^n$ of $\theta$ (suppose $\pi(\theta)$) ; when the data dist$^n$ is known to us.

## Our objective :-

To find a "good estimator" of $\pi(\theta)$ based on the random sample $(X_1, \dots, X_n)$ obtained from the data dist$^n$.

### A few fundamental questions :-

1. What is an estimator ?? What is an estimate ??
2. How to find an estimator ??
3. What does it mean by "good" estimator ??

Answer of 1. ← **Estimator :-** $T_n(X_1, \dots, X_n)$, which is a $f^n$ of $(X_1, \dots, X_n)$, is an estimator. When $X_1 = x_1, \dots, X_n = x_n$ (i.e., $\mathcal{X} = \{x_1, \dots, x_n\}$), the estimate of $T_n(X_1, \dots, X_n)$ will be $T_n(x_1, \dots, x_n)$.

Example:- Let $X_1, \dots, X_n \sim N(\mu, 1)$.

i) $T_n^{(1)}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$ is an estimator of $\mu$.

ii) $T_n^{(2)}(X_1, \dots, X_n) = \frac{X_1 + X_7}{2}$ " " " " $\mu$.

iii) $T_n^{(3)}(X_1, \dots, X_n) = \frac{X_{(3)} + X_{(109)}}{2}$ " " " " $\mu$.

iv) $T_n^{(4)}(X_1, \dots, X_n) = 0$ " " " " $\mu$.

$\forall \ i = 1, \dots, 4$, $T_n^{(i)}(X_1, \dots, X_n)$ is an estimator of $\mu$.

Suppose, $x_1 = 0.37$, $x_2 = 1.12$, $x_3 = 3.79$, $\dots$ $x_{300} = 7.52$, we can estimate $\mu$ based on each $T_n^{(i)}(X_1, \dots, X_n)$.

Answer of 2nd question: 2) There are a few well-known methods to find an estimator:-

a) Method of Moment.

b) Maximum likelihood estimation.

c) Least square estimation.

d) Least absolute deviation estimation.

Point estimation:- $T_n(X_1, \ldots, X_n)$ is estimator & for a given data $x = \{x_1, \ldots, x_n\}$ (of the realization of $(X_1, \ldots, X_n)$) estimate the unknown parameter $\theta$ pointwise.

Point estimation/estimator $\rightarrow$ Unbiased estimator

Point estimation/estimator $\rightarrow$ Consistent estimator.

**Remark:-** Unbiased estimation is a finite sample property of an estimator whereas consistent estimation is (consistency) a large sample property of an estimator.

**Unbiased estimator:-** Set up :- $X_1, \dots, X_n \sim f(x|\theta)$

p.m.f / p.d.f.    unknown parameter

$\theta \in \Theta$ parameter space.

$T_n = T_n(X_1, \dots, X_n)$ is said to be an unbiased estimator of $\theta$ if

$$E(T_n) = \theta \quad \text{for all } \theta \in \Theta.$$

Example:-

Let $X_1, \ldots, X_n \sim N(\mu, 1)$.

$$T_n^{(1)} = X_1, \quad T_n^{(2)} = \frac{X_1 + X_2}{2}, \quad T_n^{(3)} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$T_n^{(4)} = 0.7 \times T_n^{(1)} + 0.17 \times T_n^{(2)} + 0.13 \times T_n^{(3)}$$

$$T_n^{(5)} = \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{5}{n}.$$

Study:- i) $E\left[T_n^{(1)}\right] = E(X_1) = \mu$.

ii) $E\left(T_n^{(2)}\right) = E\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}\left[E X_1 + E X_2\right] = \frac{1}{2}[\mu + \mu] = \mu$.

iii) $E\left(T_n^{(3)}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$.

iv) $E\left(T_n^{(4)}\right) = 0.7\, E\left(T_n^{(1)}\right) + 0.17\, E\left(T_n^{(2)}\right) + 0.13\, E\left(T_n^{(3)}\right) = 0.7 \times \mu + 0.17\mu$
$+ 0.13\mu = \mu$.

iv) $E\left(T_n^{(5)}\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) + E\left(\frac{5}{n}\right)$

$= \frac{1}{n}\sum_{i=1}^{n} \mu + \frac{5}{n} = \mu + \frac{5}{n}.$

Remark:- $T_n^{(5)}$ is NOT an unbiased estimator but as $n \to \infty$, $E\left(T_n^{(5)}\right) = \mu$, and hence, $T_n^{(5)}$ is an asymptotically unbiased estimator.

Remark:- The construction of $T_n^{(4)}$ indicates that the u.e. ~~is not~~ may NOT be unique. (if $E(T_n^{(1)}) = \theta$ & $E(T_n^{(2)}) = \theta$, then $E\left(a_1 T_n^{(1)} + a_2 T_n^{(2)}\right) = \theta$ as long as $a_1 + a_2 = 1$.)