Sample : $x_1 \xrightarrow{} f(\theta),\ x_2 \xrightarrow{} f(\theta), \dots, x_n \xrightarrow{} f(\theta)$: $g(\theta)$

ordered sample: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

First order statistic. Highest order statistic.

$$x_{(1)} = \min\{x_1, x_2, \dots, x_n\}.$$

$$x_{(n)} = \max\{x_1, x_2, \dots, x_n\}.$$

Ordered sample: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

Median : middlemost obs of the dataset

$$\text{Median} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & n \text{ is odd.} \\ \dfrac{x_{(n/2)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & n \text{ is even.} \end{cases}$$

$O_1\ O_2\ O_3\ O_4\ O_5$

$n = 5$

$$\text{Med} = \left(\frac{n+1}{2}\right)^{th} \text{obs.}$$

$O_1\ O_2\ O_3\ O_4$

$$\frac{O_2 + O_3}{2}$$

$n = 4.$

$$\text{Med} = \frac{\left(\frac{n}{2}\right)^{th}\text{obs} + \left(\frac{n}{2}+1\right)^{th}\text{obs}}{2}$$

Med.

Dataset: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

$(1/4)^{th}$ $(1/2)$ $(3/4)^{th}$.

Dataset: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

Q_1 Q_2 Q_3.

Quartiles.

Q_1 : obs below which $(1/4)^{th}$ of the obs lie.

Q_2 : med : obs below which $(1/2)$ of the obs. lie.

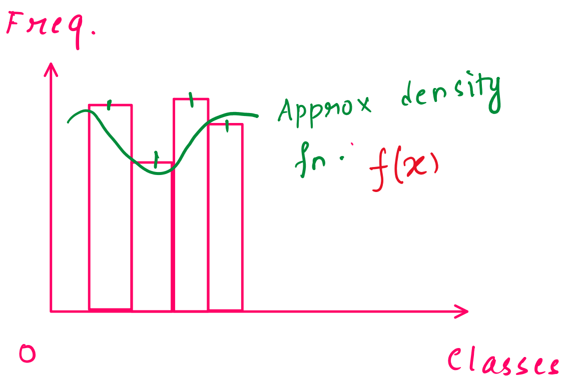
Q_3 : upper quartiles : obs below which $(3/4)^{th}$ of the obs lie.

Dataset: Values that split the dataset [of size n] into k-segments are known as "quantiles"

Eg: 80^{th} percentile : obs below which $\left(\frac{80}{100} \times n\right)$ obs lie.

Histogram

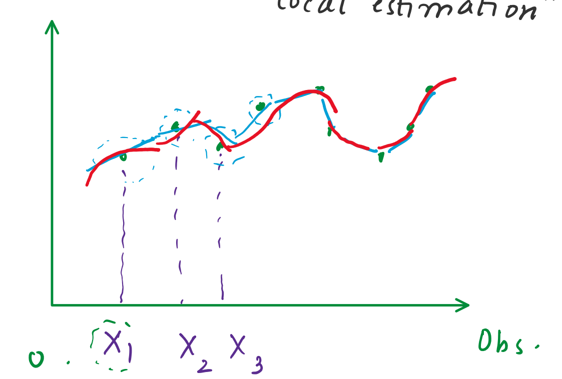
Empirical distribution



Dataset : $x_1, x_2, \ldots, x_n$.

Fix bins... group the data.

Kernel Estimation



Approx bands chosen: Kernel fns. $K(t)$.

Obj: Obtain $\hat{f}(x)$.

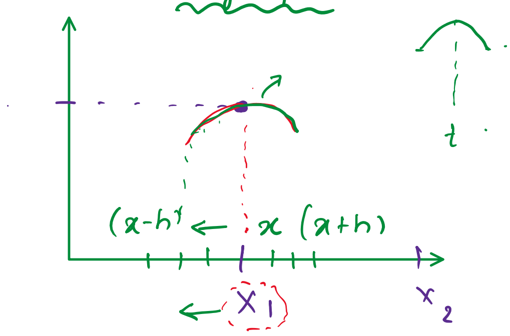
Theoretically: $f(x) = F'(x)$.

Emperical cum distn.

$$P(X \leq x) = \frac{1}{n}\{\# x_i \leq x\}.$$

$\hat{f}(x)$

Magnified



Define: $F_n(x) = \frac{1}{n}\{\# X_i \leq x\}$.

$$F'(x) \simeq \frac{F_n(x+h) - F_n(x-h)}{2h}$$

$$F'(x) \simeq \frac{F_n(x+h) - F_n(x-h)}{2h}$$

$$F_n(x+h) = \frac{1}{n}\{\# X_i \leq (x+h)\}.$$

$$F_n(x-h) = \frac{1}{n}\{\# X_i \leq (x-h)\}$$

$$\therefore \quad F_n(x+h) - F_n(x-h) = \frac{1}{n}\{\# X_i \in (x-h, x+h)\}.$$

$$\& \quad F'(x) = \frac{1}{2nh}\{\# X_i \in (x-h, x+h)\}$$

As $f(x) = F'(x)$

$$\therefore \quad \hat{f}(x) = \frac{1}{2nh}\{\# X_i \in (x-h, x+h)\}.$$

Kernel: weight fns, put on obs to obtain $\hat{f}(x)$.