

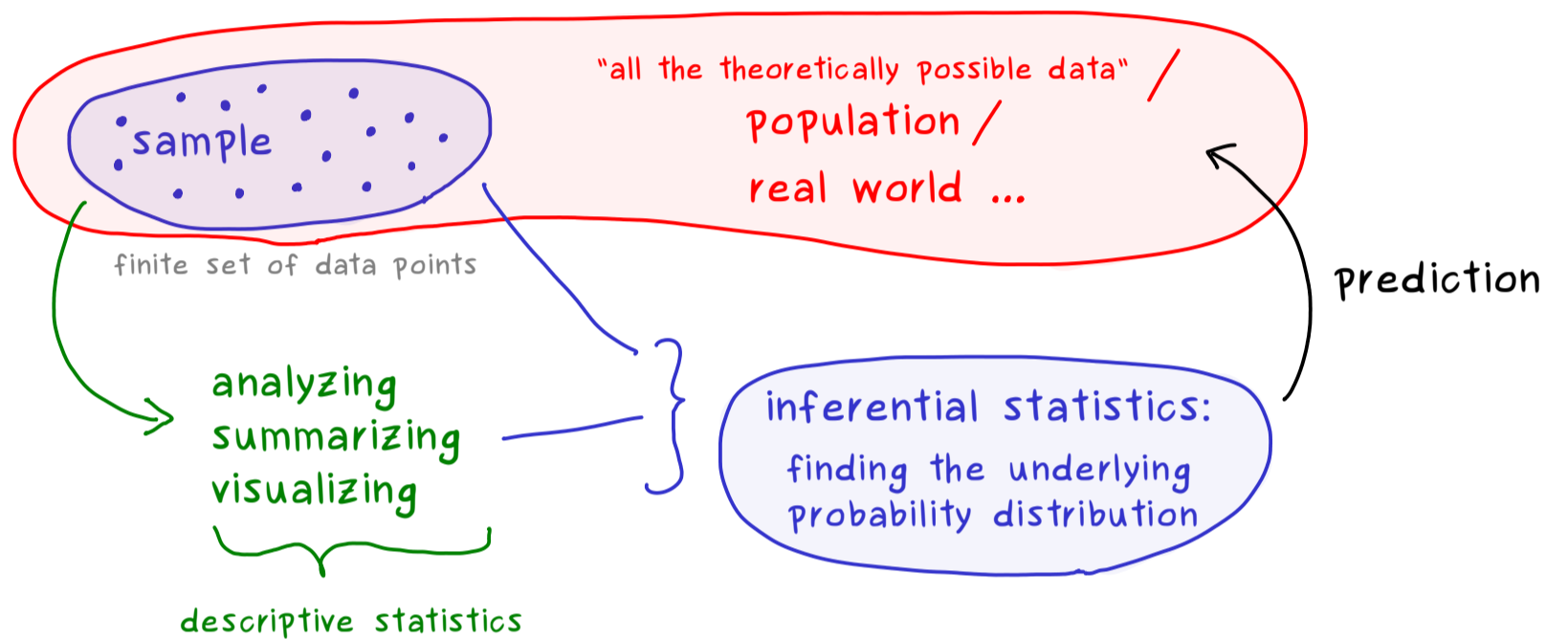
Probability Theory - Part 33

Statistics

descriptive statistics

inferential statistics

The rough idea:

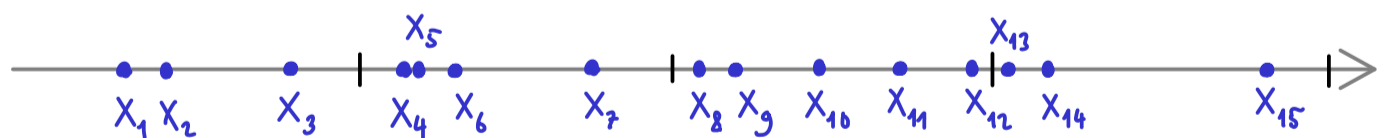


Definition: sample in \mathbb{R}^d : $(X_1, X_2, X_3, \dots, X_n)$ with $X_j \in \mathbb{R}^d$
 $n \in \mathbb{N}$

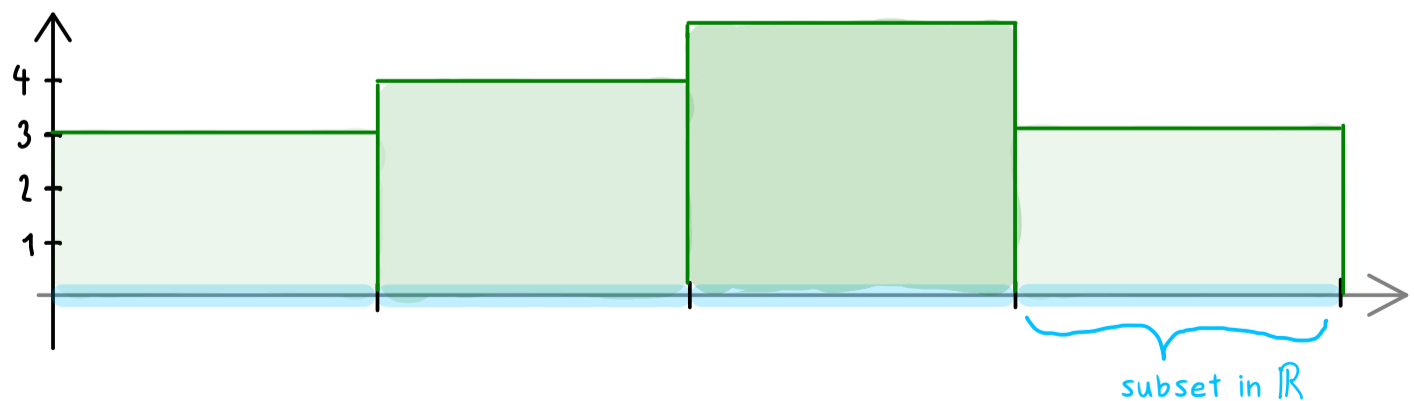
sample in \mathbb{R} : $(X_1, X_2, X_3, \dots, X_n)$ with $X_j \in \mathbb{R}$

can be ordered: $X_1 \leq X_2 \leq \dots \leq X_n$

visualization



histogram
(grouping)



Definition: For a given sample in \mathbb{R} $(X_1, X_2, X_3, \dots, X_n)$ and subset $A \subseteq \mathbb{R}$, we define:

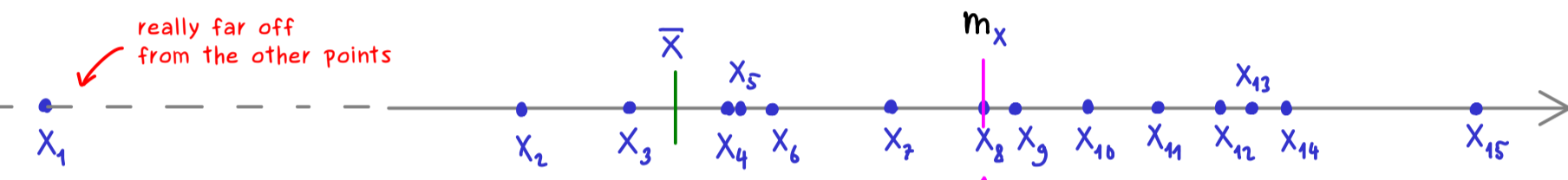
absolute frequency of A $f_{\text{abs}}(A) := \#\{k \in \mathbb{N} \mid x_k \in A\}$

relative frequency of A $f_{\text{rel}}(A) := \frac{\#\{k \in \mathbb{N} \mid x_k \in A\}}{n} \in [0, 1]$

Definition: For a given sample in \mathbb{R} $X := (X_1, X_2, X_3, \dots, X_n)$

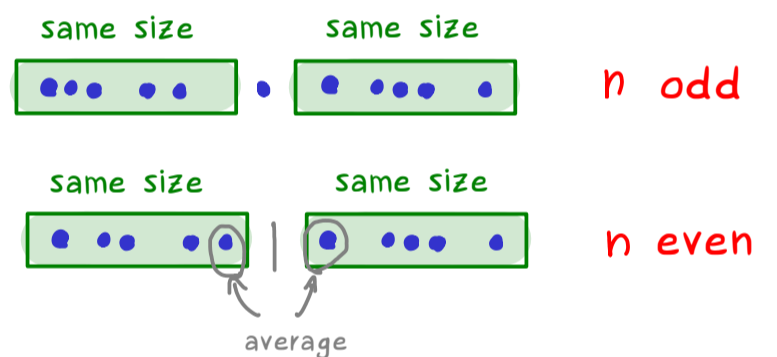
we define the sample mean:

$$\bar{X} := \frac{1}{n} \sum_{k=1}^n X_k$$



median: middle point of the sample

$$m_x := \begin{cases} X_{\frac{n+1}{2}}, & n \text{ odd} \\ \frac{1}{2}(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}), & n \text{ even} \end{cases}$$



unbiased sample variance:

$$s_x^2 := \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

↑ makes it unbiased