



## EMPIRICAL MEASURE

### Why

There is a natural probability measure on a measurable space to associate with a dataset from the base set of that space.

### Definition

The *empirical measure* for a dataset in some measurable space is the measure which associates to each event the proportion of the records which are elements of that event.

### Notation

Let  $(a^1, \dots, a^n)$  be a dataset in a measurable space  $(A, \mathcal{A})$ . Let  $P : \mathcal{P}(A) \rightarrow [0, 1]$  be the probability measure that assigns to each set  $B \subset A$  the number

$$P(B) = \frac{1}{n} \left| \{k \in \{1, \dots, n\} \mid a^k \in B\} \right|.$$

Then  $P$  is the empirical measure.



