Empirical Distribution

1 Why

A natural distribution to associate with a dataset is to assign probabilities which are the proporitions.

2 Definition

The *empirical distribution* of a dataset is to assign probabilities to outcomes in accordance to their proportions of appearance in the dataset. The proporitions are nonnegative and sum to one, so this definition yields in a probability distribution.

2.1 Notation

Let $(a^1, \ldots, a^n)$ be a data set in A. Let $q : A \to \mathbf{R}$ be defined by

$$q(a) = \frac{1}{n} \left| \left\{ k \in \{1, \ldots, n\} \;\middle|\; a^k = a \right\} \right|$$

Then q is the empirical distribution of $(a^1, \ldots, a^n)$.