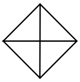


Why

A natural distribution to associate with a dataset is to assign to each outcome a probability which reflects the number of times it appears in the dataset.

Definition

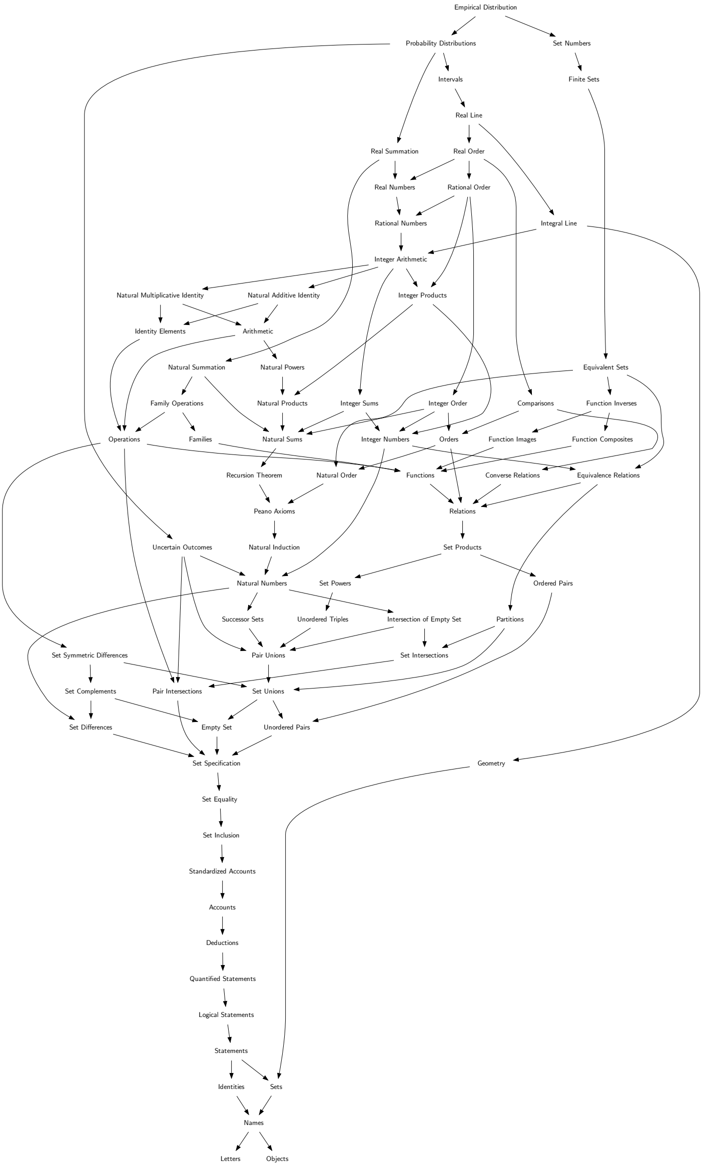
The *empirical distribution* of a dataset is the function which associates to each outcome the proportion times it appears in the dataset. Since the proporitions are nonnegative and sum to one, the function is a probability distribution.

Notation

Let A be a non-empty set and $(a^1, \dots, a^n)$ be a dataset in A. The empirical distribution $q : A \to \mathbf{R}$ of $(a^1, \dots, a^n)$ is

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

In other words, to give the probability of outcome $a \in A$, we count the number of times it appeared in the dataset of size n, and then divide by n.


Empirical Distribution
Probability Distributions
Set Numbers
Intervals
Finite Sets
Real Line
Real Summation
Real Order
Real Numbers
Rational Order
Rational Numbers
Integral Line
Integer Arithmetic
Natural Multiplicative Identity
Natural Additive Identity
Integer Products
Identity Elements
Arithmetic
Natural Summation
Natural Powers
Equivalent Sets
Family Operations
Natural Products
Integer Sums
Integer Order
Comparisons
Function Inverses
Operations
Families
Natural Sums
Integer Numbers
Orders
Function Images
Function Composites
Recursion Theorem
Natural Order
Functions
Converse Relations
Equivalence Relations
Peano Axioms
Relations
Uncertain Outcomes
Natural Induction
Set Products
Natural Numbers
Set Powers
Ordered Pairs
Successor Sets
Unordered Triples
Intersection of Empty Set
Partitions
Set Symmetric Differences
Pair Unions
Set Intersections
Set Complements
Pair Intersections
Set Unions
Set Differences
Empty Set
Unordered Pairs
Set Specification
Geometry
Set Equality
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Identities
Sets
Names
Letters
Objects