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

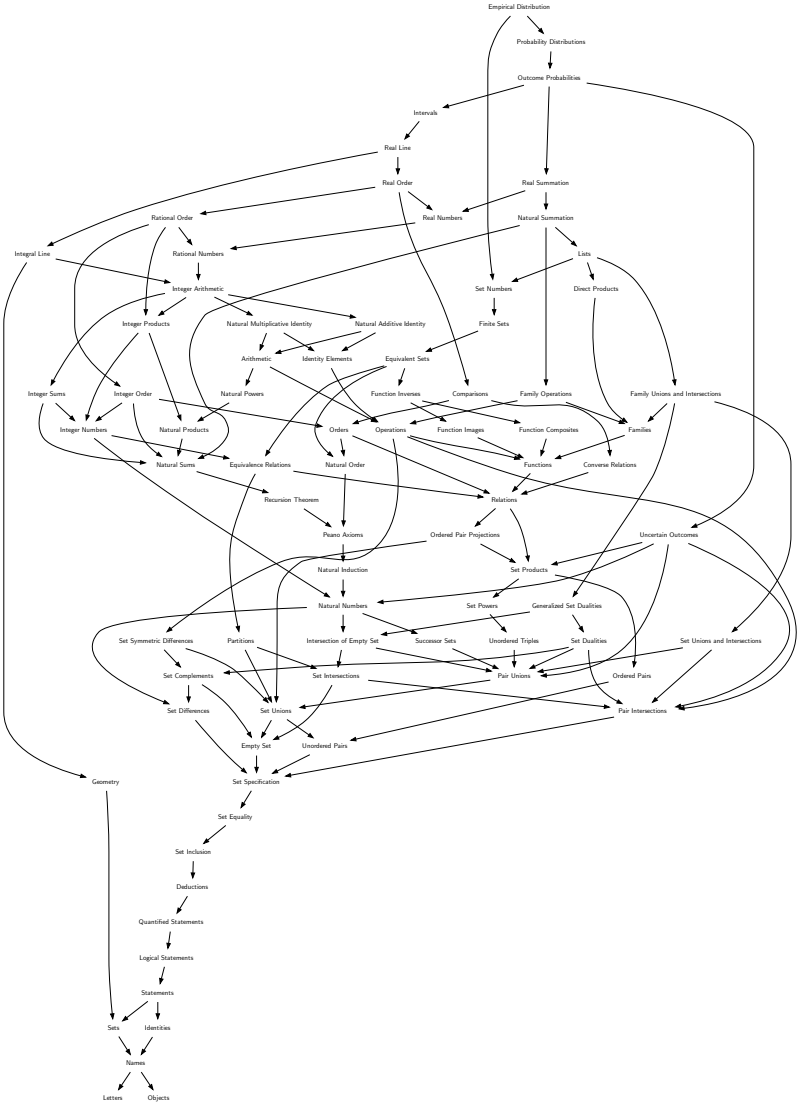
The *empirical distribution* of a dataset is the function which associates to each outcome the proportion times it appears in the dataset. Since the proporitions are nonnegative and sum to one, the function is a probability distribution.

Notation

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

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

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