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

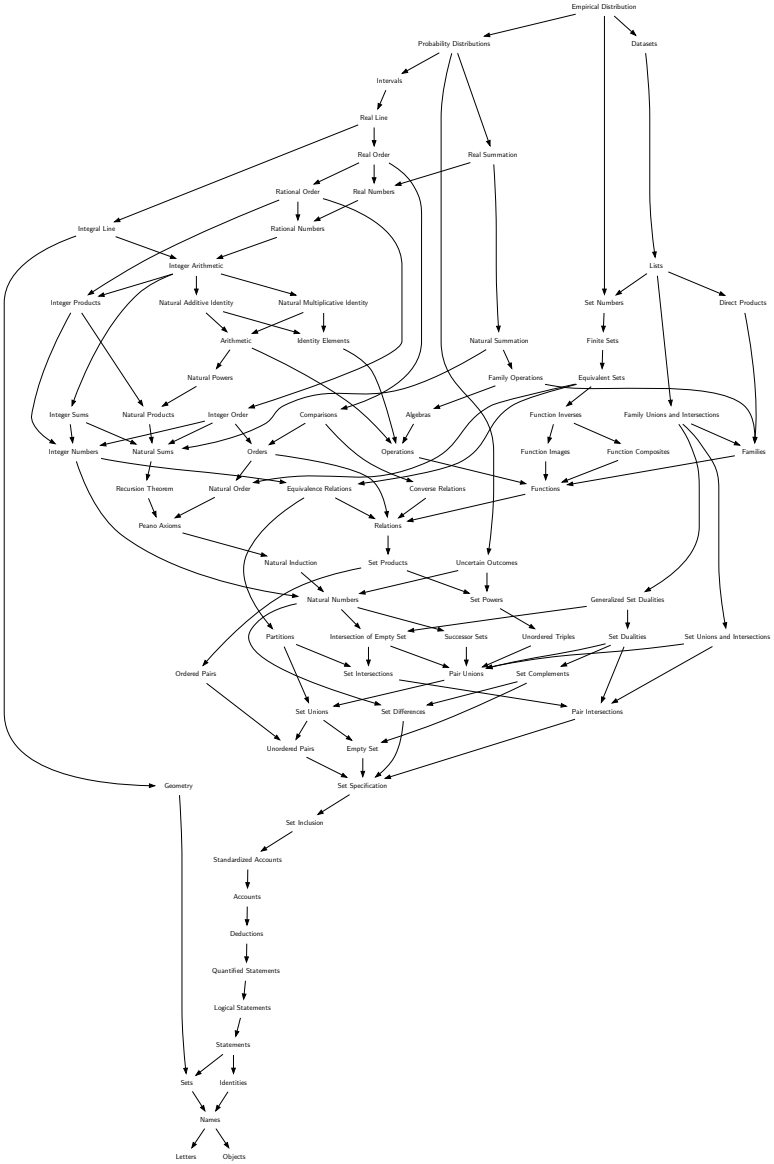
The *empirical distribution* of a dataset is the function which associates to each outcome the proportion times it appears in the dataset. Since the proporitions are nonnegative and sum to one, the function is a probability distribution.

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 defined by

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

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