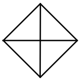


Why

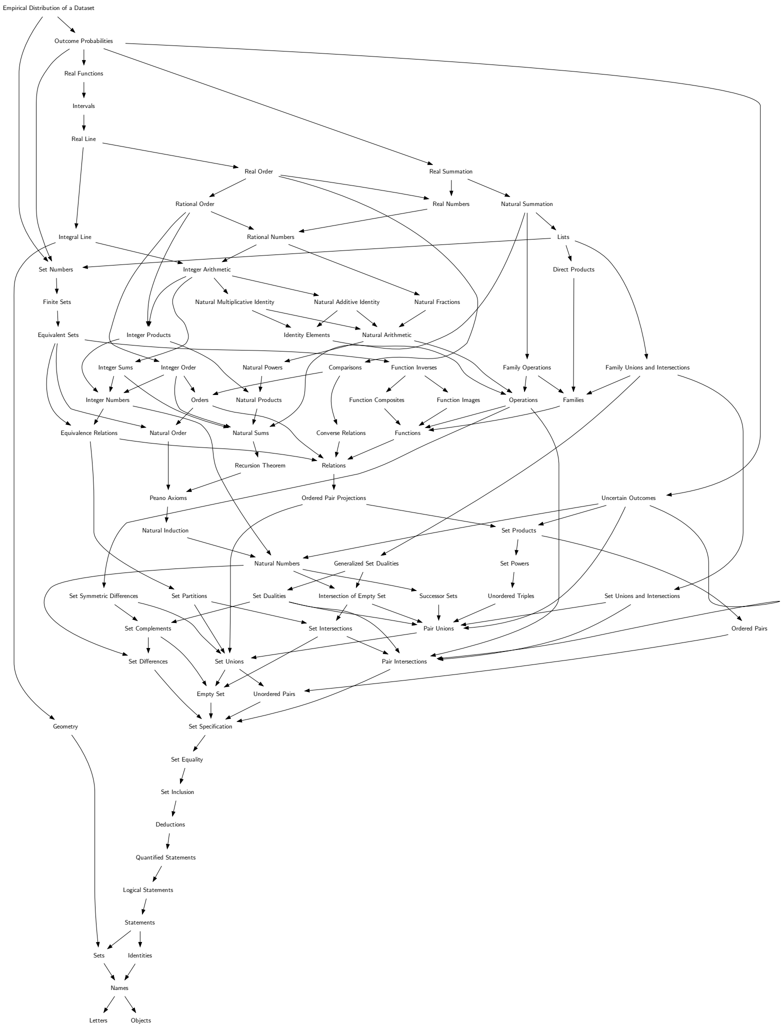
A natural distribution to associate with a dataset is to assign to each outcome a probability which reflects the number of times it appears in the dataset.

Definition

Given a dataset $x_1, \dots, x_n$ is a finite set X, the *empirical distribution* is the function $q : X \to \mathbf{R}$ which associates each outcome with the proportion of times it appears in the dataset. In other words, q is defined by

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

The function q is clearly a distribution, since the proportions are nonnegative and sum to one.


Empirical Distribution of a Dataset
Outcome Probabilities
Real Functions
Intervals
Real Line
Real Order
Real Summation
Rational Order
Real Numbers
Natural Summation
Integral Line
Rational Numbers
Lists
Set Numbers
Integer Arithmetic
Direct Products
Finite Sets
Natural Multiplicative Identity
Natural Additive Identity
Natural Fractions
Equivalent Sets
Integer Products
Identity Elements
Natural Arithmetic
Integer Sums
Integer Order
Natural Powers
Comparisons
Function Inverses
Family Operations
Family Unions and Intersections
Integer Numbers
Orders
Natural Products
Function Composites
Function Images
Operations
Families
Equivalence Relations
Natural Order
Natural Sums
Converse Relations
Functions
Recursion Theorem
Relations
Peano Axioms
Ordered Pair Projections
Uncertain Outcomes
Natural Induction
Set Products
Natural Numbers
Generalized Set Dualities
Set Powers
Set Symmetric Differences
Set Partitions
Set Dualities
Intersection of Empty Set
Successor Sets
Unordered Triples
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