Why

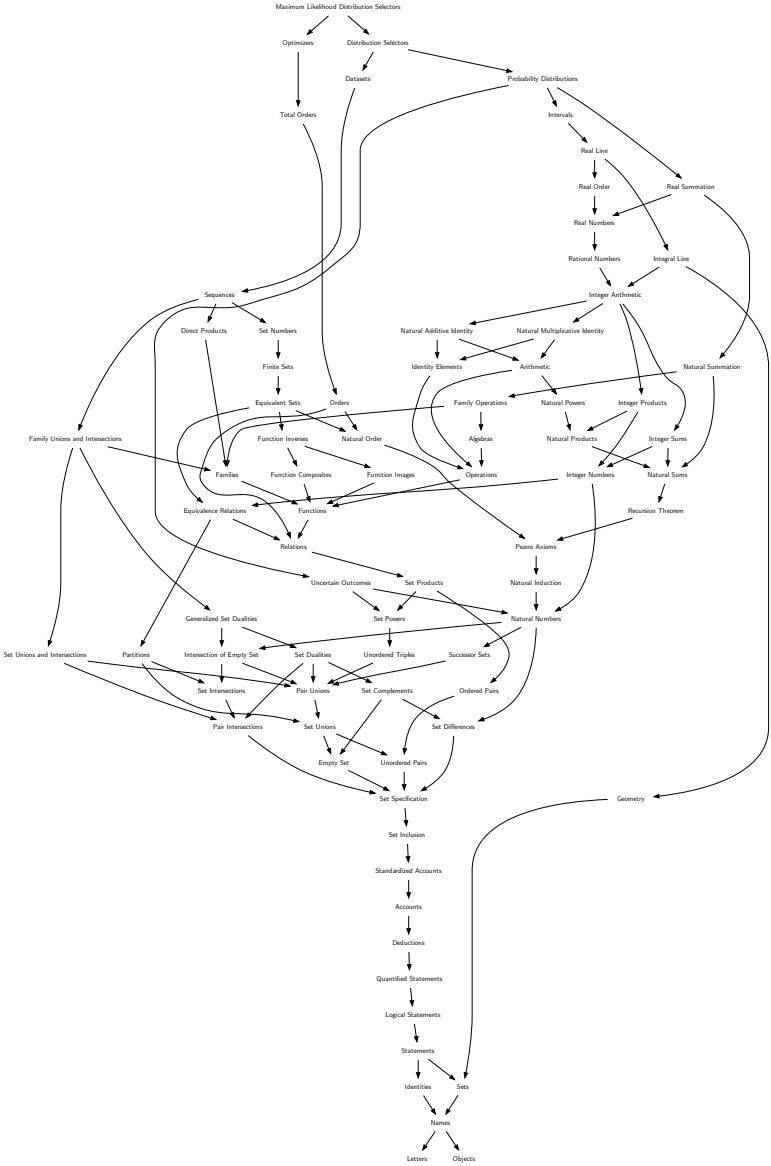
We provide a principle for distribution selection based on maximizing the probability of the dataset.

Definition

We have a set of outcomes A and a dataset $(a^1, \ldots, a^n)$. We want a distribution $p : A \to \mathbf{R}$. We define $\bar{p} : A^n \to \mathbf{R}$ by $\bar{p}(a^1, \ldots, a^n) = \prod_{i=1}^n p(a^i)$.

The *principle of maximum likelihood* says to solve:

$$\begin{aligned} \textbf{find} \quad & p, \text{ a distribution} \\ \textbf{to maximize} \quad & \bar{p}(a^1, \ldots, a^n) \end{aligned}$$


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