Why

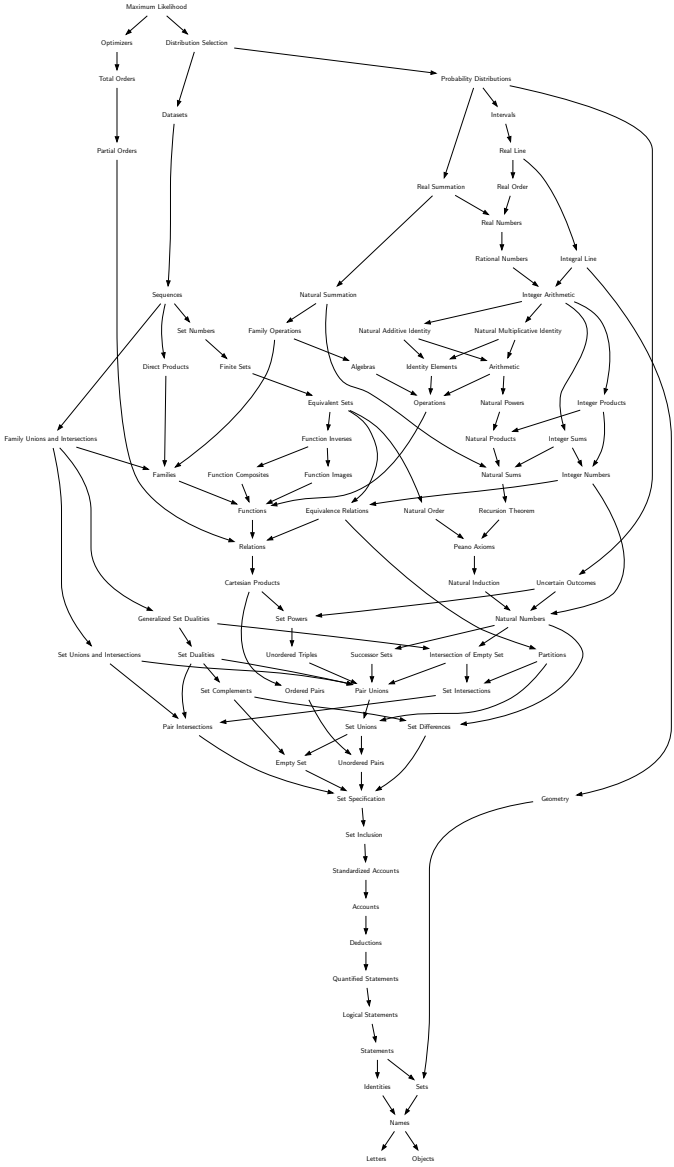
We provide a principle for distribution selection based on maximizing the probability of the dataset.

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
Optimizers
Distribution Selection
Total Orders
Datasets
Probability Distributions
Intervals
Partial Orders
Real Line
Real Summation
Real Order
Real Numbers
Rational Numbers
Integral Line
Sequences
Natural Summation
Integer Arithmetic
Set Numbers
Family Operations
Natural Additive Identity
Natural Multiplicative Identity
Direct Products
Finite Sets
Algebras
Identity Elements
Arithmetic
Equivalent Sets
Operations
Natural Powers
Integer Products
Family Unions and Intersections
Function Inverses
Natural Products
Integer Sums
Families
Function Composites
Function Images
Natural Sums
Integer Numbers
Functions
Equivalence Relations
Natural Order
Recursion Theorem
Relations
Peano Axioms
Cartesian Products
Natural Induction
Uncertain Outcomes
Generalized Set Dualities
Set Powers
Natural Numbers
Set Unions and Intersections
Set Dualities
Unordered Triples
Successor Sets
Intersection of Empty Set
Partitions
Set Complements
Ordered Pairs
Pair Unions
Set Intersections
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