Why

Let $X = \{a, b\}$ and $Y = \{0, 1\}$. Define $f : X \to Y$ by $f \equiv 0$.

Consider "selecting" elements of X so to determine f. Since f is functional, the program is straightforward. Observe f at a and at b, and then any consistent inductor will determine f.

So, in some sense, the dataset $((a, 0), (b, 0))$ is "good" in that "reasonable" learning algorithms (those which are consistent) will determine f from a dataset consistent with f. What of more general domains X?

Definition

Let X and Y be sets. Let $\{G_n : (X \times Y)^n \to (X \to Y)\}_{n \in \mathbb{N}}$ be a family of inductors.

An *input design* (or *experiment design* or *experimental design*) of size n is a sequence of inputs $x_1, \ldots, x_n \in X$. The interpretation is that we will observe a dataset $(x_1, y_1), \ldots, (x_n, y_n)$ consistent with some function and that we will "infer" the function $G((x_i, y_i)_{i=1}^n)$.

Some authors define a design as a sequence $x_1, \ldots, x_\ell$ and nonzero integers $n_1, \ldots, n_\ell$ so that $\sum_{i \leq \ell} n_i = n$. The correspondence with our definition is obvious.

A *conditional input design* (of size k) for a dataset $(x_1, y_1), \ldots, (x_m, y_m)$ is a design $x_{m+1}, \ldots, x_{m+k} \in X$. The interpretation is that we will observe a dataset $(x_1, y_1), \ldots, (x_{m+k}, y_{m+k})$ consistent with some function and that we will "infer" the function $G((x_i, y_i)_{i=1}^{m+k})$.

For both of these designs, our interpretation of these objects is "experimental." We call $x \in X$ the *experimental conditions* and we call X the *experimental domain*. We think of "choosing" a design and then "running an experiment" or "collecting data." The experimental conditions are understood to be freely chosen (within the domain X). We call an element of the dataset y_i an *observed response* or *yield*.

Parametrically consistent datasets

Let $f : X \to Y$ and suppose the dataset is consistent with f. Let Θ be a nonempty set. Suppose there exists $\theta^\star \in \Theta$ and $g : \Theta \times X \to Y$ so that $g(\theta^*, x) = f(x)$ for all $x \in X$ (and so denote f by $g_{\theta^\star} : X \to Y$).

In this case we call θ the *parameters* or *parameter system* and we call Θ the *parameter domain* (or *parameter space*). We call the dataset *parametrically consistent.*

This decomposition (θ, x) is interpreted as reflecting what is within and without of the experimenter's control. The conditions can be chosen by the experimenters, but the parameters are fixed, or "chosen by nature." The experimenter "controls" the conditions and nature "controls" the parameters.

Input Designs

Consistent Inductors

Predictors

Lists

Direct Products   Set Numbers

                         Finite Sets

                         Equivalent Sets

Family Unions and Intersections   Natural Order   Function Inverses

                    Families   Function Composites   Function Images

                         Peano Axioms   Functions   Equivalence Relations

                         Natural Induction   Relations

                                                           Set Products

Generalized Set Dualities   Natural Numbers   Set Powers

Set Dualities   Set Unions and Intersections   Intersection of Empty Set   Successor Sets   Unordered Triples   Partitions

Set Complements   Pair Unions   Set Intersections   Ordered Pairs

Set Differences   Pair Intersections   Set Unions

Empty Set   Unordered Pairs

Set Specification

Set Equality

Set Inclusion

Deductions

Quantified Statements

Logical Statements

Statements

Identities   Sets

Names

Letters   Objects

4