Pure Data Foundations of Mathematics

Saul Youssef
Department of Physics
Boston University
youssef@bu.edu

April 13, 2023

Abstract

This is an abstract.

1 Introduction

2 Foundation

Any system of reasoning must necessarily have at least one foundational concept which is understood even before a first definition. In our case, the one foundational concept is the **finite sequence**. We assume that finite sequences and obvious operations on finite sequences are understood. On the other hand, logic, logical values and predicates (foundational in ZFC) and types (foundational in dependent type theory) are not assumed. These will be defined in terms of finite sequences as we proceed. Given finite sequences, we can define *data* as follows:

- **data** is a finite sequence of **codas**, and

- a **coda** is a pair of **data**.

The pairing of data in the definition of a coda is indicated by a colon, so (), (:) and $((:) : (:)(:)(: (:)))$ are examples of valid data. We refer to this a "pure data" since it is "data made of nothing." As we shall see, variables, functions, definitions, language expressions, logical values, morphisms, categories, types, propositions, theorems and other mathematical objects all appear as pure data of different varieties. Data in general has a small natural algebra with two binary operations:

$A\ B$ is the concatenation of data A and data B as finite sequences, and

$A : B$ is the pairing of data A and data B into a single coda.

Without parenthesis, the colon binds right-first and is weaker than concatenation, so for instance, $A : B : C$ means $(A : (B : C))$ and $A : B\ C$ means $(A : (B\ C))$.

A *context* in coda is a partial function from coda to data. A context which is deemed to be *valid* defines an equivalence relation "=" on data, via assuming relations

$A\ B = \delta(A)\ B = A\ \delta(B)$

$A : B = \delta(A) : B = A : \delta(B)$

for any data A and B. Here δ is the valid context in question, extended from a partial function from coda to data into to a total function from data to data by adding identities. In other words, relations are assumed, so as to force δ to be the same as the identity function on data. This equality, in turn, defines the answers to mathematical questions, as we will see. The issue of which contexts are valid is fixed by the following.

AXIOM OF DEFINITION. *The empty context is a valid context. If δ is a valid context with corresponding equality $=$, and if d is a partial function from coda to data with $d(c) = d(c')$ if $c = c'$ in δ, then $\delta \cup d$ is also a valid context.*

A partial function from coda to data, such as d in the axiom above, is called a *definition*. The axiom of definition defines what constitutes a valid definition. The axiom of definition is the one and only axiom of coda.

If data A is in the domain of context δ and if $\delta(A) = A$ and if $|A| = |B| = 1$ for any $A = B$, then data A is said to be an **atom**. Note that if A is an atom, A remains an atom independent of any new definitions added to δ via the Axiom of definition. If data has one or more atoms in it's sequence, it is said to be **atomic** data. In this sense, atoms cannot be destroyed by future definitions. This makes it possible to define, for instance, a stable 0-bit, 1-bit, stable bit sequences and stable byte sequences and many other types of storage as indicated in figure 1.

3   Logic

Within coda, items with mathematical meaning are merely pure data of different kinds. This means, roughly speaking, that the answer to mathematical questions will be of the form

Is data A equal to data B?

Since equality is also a definition in context, the answer to the question $A = B$ is also data. But if everything about $A = B$ is encoded in the concrete data $(= A : B)$, this specific data should include any logical interpretation of $A = B$. This suggests that "logic" in should be identified as the coarsest non-trivial classification of data in general. A glance at any pure data suggests that there is such a classification since pure data is "made from () and (:)." It is clear that if there was a definition

$(:) \rightarrow ()$

then all data collapse and be equal to the empty data. On the other hand, the alternative

$(:) \rightarrow (:)$

makes (:) into an atom which cannot be modified by other definitions. This suggests that logic should have at least two values and () and (:) should have different logical values. This motivates the definitions

– data A is **true** if A is equal to the empty data.

– data A is **false** if A contains an atom in it's sequence of codas.

– data A is **undecided** otherwise.

This logic could be thought of as "not quite two valued" where undefined data like (foo : bar) represents data which is currently undefined, but which may receive a value if more definitions are added to the current context. The stability of true or false data under new definitions makes these useful concepts. We use "always" to refer to future definitions as in

– True data is "always true."

– False data is "always false."

Undecided data, on the other hand, may become true or false, depending upon added definitions. Also, as we will see, there is some undecided data which can never become true or false with any choice of future definitions. Such data is called **undecidable**. This will be discussed further in section X.

Although the logic of coda is not the same as classical logic, we will argue that this is the "correct" logic for reasoning in general. It is also quite close to classical logic in the sense that there is a definition that corresponds to each of the familiar 16 binary operations of classical propositional logic. For instance, when data A and B are either true or false, the data $(XOR\ A : B)$ has the value () (for true) and (:) for false depending on the standard truth table values for XOR. If either A or B are undecided, no definition applies to $(XOR\ A : B)$ and it "waits for A and B to be resolved in future definitions."

4   Language

5   Proof and Computation

6   Spaces

7   Is Mathematics consistent?

8   Mathematical Machine Learning

9   Summary

References

[1] R.T.Cox, Am.J.Phys. 14, 1 (1946).

[2] S.Youssef, Mod.Phys.Lett A9, 2571 (1994).

[3] S.Youssef, Phys.Lett. A204, 181(1995).

[4] S.Youssef, Mod.Phys.Lett A6, 225-236 (1991).

[5] S.Youssef in: proceedings of the Fifteenth International Workshop on Maximum Entropy and Bayesian Methods, ed. K.M.Hanson and R.N.Silver, Santa Fe, August(1995).

[6] S.K.Srinivasan and E.C.G.Sudarshan, J.Phys.A. Math Gen.27(1994).

[7] S.K.Srinivasan, J.Phys.A (23) 8297 (1997).

[8] S.K.Srinivasan, J.Phys.A Math.Gen. 31(1998).

[9] P.A.M. Dirac, Proc.Roy.Soc.Lond. A 180(1942).

[10] W.Muckenheim et al., Phys.Rep. 133, 339(1983).

[11] S.Gudder, J.Math.Phys.29, 9(1988); Found.Phys.19, 949(1989); Int.Journ.Theor.Phys.,31(1992)15.

[12] R.P.Feynman in: *Quantum Implications*, eds. B.J.Hiley and F.David Peat(Routledge and Kegan Paul, 1987).

[13] Y.Tikochinsky, Int.J.Theor.Phys. 27, 543(1988); J.Math.Phys. 29 (1988).

[14] F.H.Frohner, in: Maximum Entropy and Bayesian Methods, ed: A.Mohammad-Djafari and G.Demoments, Kluwer Academic Publishers, (1993).

[15] A.Caticha, Phys.Rev.A57, 1572 (1998).

[16] A.Steinberg, Phys.Rev.Lett 13, 2405 (1995); Phys.Rev. A52, 32 (1996).

[17] A.V.Belinskii, JETP letters, 69, 301(1994).

[18] D.J.Miller, Phys.Lett. A (1996).

[19] W.Muckenheim, Phys. Lett. A 175 (1993).

[20] A.Krennikov, Phys.Lett. A 200(1995).

[21] I.Pitowsky, Phys.Rev.Lett. 48, 1299 (1982); Phys.Rev.D27, 2316(1983).

[22] E.T.Jaynes, "Bayesian Methods: General Background," in *The Fourth Annual Workshop on Bayesian/Maximum Entropy Methods in Geophysical Inverse Problems*, ed: J.H.Justice, Cambridge University Press (1985).

[23] E.T.Jaynes, "Clearing up Mysteries–The Original Goal," in *Maximum–Entropy and Bayesian Methods*, ed: J.Skilling, Kluwer, (1989).

[24] E.T.Jaynes, "Probability Theory as Logic," in: *Maximum–Entropy and Bayesian Methods*, ed: P.F.Fougere, Kluwer, (1990).

[25] E.T.Jaynes, *Probability Theory: The Logic of Science*, `http: bayes.wustl.edu/etj/etj.html`.

[26] Assuming that $X \times T$ is a sublattice of L implies that probabilities such as $(a \to N_t)$ can occur where N is a non–measurable set. This does not cause problems because we only assume that $(a \to b \vee c) = (a \to b) + (a \to c)$ for pairs (b, c) and not for countable subsets $b_1, b_2, \ldots$ with $b_i \wedge b_j = 0$ for all $i \neq j$. "Measure space" in this paper refers to a measure space with a finite measure.

[27] Hurwitz theorem, see *Spinors and Calibrations* by F.Reese Harvey, Academic Press, (1990).

[28] J.S.Bell, Physics, 1(1964) 195; J.S.Bell, Rev.Mod.Phys. 38 (1966) 447. See N.D.Mermin, Rev.Mod.Phys. 65(1993) 803 for a review.

[29] J.J.Sakurai, *Modern Quantum Mechanics*, Addison–Wesley (1994).

[30] S.R.Wilkinson et al., Nature, 387 (1997) 575.

[31] H.Risken, *The Fokker–Planck equation*, Springer-Verlag (1984).

[32] S.Lang, *Real and Functional Analysis*, Springer (1993).