Modeling Uncertain Self-Referential Semantics with Infinite-Order Probabilities

Ben Goertzel
Novamente LLC

January 9, 2008

1 Introduction

The notion of a "second-order probability" (a probability distribution defined over a space of probability distributions), while not exactly commonplace, has become less obscure in recent decades due to the advent of imprecise probability theory as developed by Peter Walley [1] and others [2]. It is straightforward to generalize the notion further to third-order probabilities and so forth; and the reason these notions are rarely discussed is not so much that there are mathematical difficulties, but rather that no one has seen much use for them.

However, so far as we have been able to tell, the notion of *infinite-order probabilities* has not been previously discussed. Intuitively, the notion of an infinite-order probability is simply "a probability distribution of probability distributions of probability distributions of ... of probability distributions." The goal of this paper is to briefly introduce these novel mathematical objects; and then to explain their apparent importance in the domain of semantics, via their ability to synthesize uncertain semantics with self-referential semantics.

2 Infinite-Order Probabilities as Hypersets

An infinite-order probability distribution (or ipdf, for short) is defined as a set S of elements, each one of which is a function $F : S \to [0, 1]$ with the property that $\sum_{s \in S} F(s) = 1$ (if S is finite, resulting in a "discrete ipdf"), or $\int_{s \in S} F(s)ds = 1$ (if S is infinite, resulting in a "continuous ipdf").

So, in short, an ipdf is a probability distribution over ipdf's. This violates the axiom of foundation and hence is not allowed in ordinary Zermelo-Frankel set theory; but it is permitted in variant set theories, for instance if one invokes the Anti-Foundation Axiom (AFA). In Section 5 below, we use the Solution Lemma stated in [4] to show that discrete ipdf's exist in set theory under the AFA. It is not hard to see that some continuous ipdf's may also exist according to the AFA.

3 Mean-Indexed Infinite-Order Probabilities

For the rest of the paper we will restrict attention to a particular subset of the class of discrete ipdf's, which we call *mean-indexed* ipdf's. These are by no means the only ipdf's of interest, but they are the only kind we have explored in depth so far. A mean-indexed ipdf (or mipdf for short) is an ipdf for which there is a linear space X so that each each element F

- has a value μ_F, drawn from X, called the "mean" associated with it
- has an ordinary, first-order pdf over S associated with it

If the elements of S are indexed F_i, the means may be denoted μ_i and the associated first-order pdf's may be denoted f_i.

In a mean-indexed ipdf, each ipdf element is essentially defined as a first-order pdf over the space of ipdf elements. This sort of ipdf is particularly easy to grasp onto, allowing for the convenient performance of practical numerical calculations.

Note that in the above characterization, there is no connection between μ_i and f_i except that both are associated with F_i. This obviously doesn't make much sense. The connection is supplied by the notion of the *consistency* of a mean-indexed ipdf, which we now define. Given a discrete mipdf S, the totality of first-order pdf's f_i in S may be represented as a weight matrix $W = \{w_{ij}\}$, with the property that each row sums to 1. The *implicit mean* of an element F_i may then be defined as $\nu_i = \sum_i w_{ij}\mu_j$. I.e., the mean of an ipdf element F_i is a weighted sum of the means of the other ipdf elements F_j, where the weighting is given by the coefficients associated with F_i. We then define a discrete mipdf S to be *mean-consistent* iff for all elements F_i of S, $\nu_i = \mu_i$. I.e., this means that for each element, the implicit mean equals the actual mean.

For any given weight matrix W, a consistent mipdf associated with W may be obtained by setting $\mu_i = v_i$ where v is the vector that satisfies $v = vW$. That is, v must be a left eigenvector of W corresponding to the eigenvalue 1. The existence of such a vector is guaranteed by the Perron-Frobenius Theorem [5].

4 Constructing MIPDF's with the Probabilistic Solution Lemma

It seems an attractive idea to generalize Aczel's [4] Solution Lemma, a fundamental result in hyperset theory, to address ipdf's rather than crisp solutions to hyperset equations. In this section we will develop a *Finite Probabilistic Solution Lemma*, along these lines, restricted to hypersets associated with (to use a terminology to be introduced just below) finite apg's. The proof of this is a combination of AFA set theory and Markov matrix theory. Extension to more general hypersets will be left for another paper.

First we review some basic concepts of AFA set theory, which may be presented in simple graph-theoretic terms. A digraph (G, E) consists of a set G of entities called "nodes," and a set E of ordered pairs of nodes, these pairs being called edges. The most common examples of graphs are finite graphs; however, the concept of an infinite graph presents no difficulties. If (n, m) is an element of E, I will write $n \to m$, and call m the child of n, and n the parent of m. Fix a set A of tags. Then a tagged digraph (G, E, t) is a digraph together with a function t that assigns a tag drawn from A to each childless node of G. Next, define an accessible pointed graph (apg) (G, E, t, p) to consist of a tagged digraph together with a distinguished node p which has the property that every node can be reached by some finite path from p. And define a decoration of an apg as a set-valued function d with domain G, satisfying $d(n) = t(n)$ if n is childless and $d(n) = \{d(m) : n \to m\}$ otherwise. That is, a decoration assigns to each childless node its tag, and to each parent node n those nodes m which are its children. Finally, let us say that an apg pictures a set b if there is a decoration d of the graph so that $d(p) = b$; that is, so that b is the set which decorates the distinguished node. Aczel's Anti-Foundation Axiom (AFA), whose formal definition in terms of bisimulations is moderately complex and will not be given here, has the implication that *Every apg pictures a unique set.* According to AFA, then, all the sets of standard set theory are still sets, but there are other sets too. Anything which is a set according to this definition, but not the classical definition is a hyperset.

There are multiple variants of non-foundational set theory, but here we will we work solely in $ZFCU - +AFA$, defined as Zermelo Fraenkel set theory with choice and urelements ("atomic elements"), and the anti-foundation axiom AFA replacing the axiom of foundation. Given this set-up, Aczel proves the *Solution Lemma*, which states roughly that for any system of equations in indeterminates $x, y, z, ...$, say

$$x = a(x, y, ...)$$
$$y = b(x, y, ...)$$
$$z = c(x, y, ...)$$
$$...$$

where each indeterminate appears exactly once in the left-hand side of an equation, then the system has a unique solution in the universe of hypersets.

The Solution Lemma may be more precisely formalized as follows. Given a collection U of urelements, we will write V^U for the hyperuniverse of sets with urelements from U . Formally, we regard a collection of set indeterminates X as extra urelements, and write $V^U[X]$ for $V^U \cup X$. By an equation in X , we mean an expression of the form $x = a$, where $x \in X$, $a \in V^U[X]$. By a system of equations in X we mean a family of equations $\{x = ax | x \in X\}$, exactly one equation for each indeterminate $x \in X$. By an assignment for X in V^U we mean a function $f : X \to V$ which assigns an element $f(x)$ of V to each indeterminate $x \in X$. Any such assignment extends in a natural way to a function $f^* : V^U[X] \to V^U$. Thus, given some $a \in V^U[X]$, one works with a canonical graph depicting a, replacing any childless nodes tagged by an indeterminate $x \in X$ with a graph depicting the set $f(x)$. Typically, we write $a(x, y, ...)$ for $f^*(a)$, so that an assignment f is a solution of an equation $x = a(x, y, ...)$ if $f^*(x) = a(f(x), f(y), ...)$. The Solution Lemma then states that every system of equations in a collection of indeterminates over V has a unique solution.

We now explore the generalization of the Solution Lemma to encompass mipdf's. For simplicity we restrict attention to the case of *finite equation systems*, which we define as systems of equations that contain a finite set of equations, each of finite size. It is easy to see that a finite equation system corresponds to a set of finite apg's, one per equation. Furthermore, it is useful to define a *flat* system of equations, which is one in which each equation is of the form

$$x_i = \{x_{i1}, x_{i2}, ...\}$$

Any finite equation system may straightforwardly be transformed into an equivalent finite flat equation system, via the introduction of additional indeterminates. For instance

$$x = \{Calvin, \{y\}\}$$
$$y = \{Hobbs, x\}$$

may be flattened by replacing the first equation with the two equations

$$x = \{Calvin, z\}$$
$$z = \{y\}$$

Next we introduce the notion of a *probabilistically tagged flat finite equation system* or *ptff* equation system, which is an flat equation system in which on the rhs of each equation, each *instance* of each ur-element, and each instance of each indeterminate, is associated with a probability value; and the probabilities used as tags on the rhs side of each individual equation, sum to 1. For instance, an example of a ptff equation system would be

$$x = \{(Calvin, .8), (z, .2)\}$$
$$y = \{(Hobbs, .5), (x, .5)\}$$
$$z = \{(y, 1)\}$$

Any ptff equation system may be associated with an $n \times n$ stochastic matrix W, where n is the sum of the number of equations and the number of urelements occurring in the equations. If the row i corresponds to an urelement then $W_{ii} = 1$ and the rest of the i'th row is 0). So for instance the matrix corresponding to the above system of equations is (using the ordering $(x, y, z, Calvin, Hobbs)$:

$$
\begin{array}{ccccc}
0 & 0 & .2 & .8 & 0 \\
.5 & 0 & 0 & 0 & .5 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{array}
$$

Suppose we have a ptff equation system PE, corresponding to a standard finite equation system E. We then define a solution to PE as the combination of a solution to E (in the standard sense given above), and an assignment of probability values to each indeterminate in E, in a way that agrees with the stationary probability vector of the stochastic matrix W corresponding to PE. Then by combining the Solution Lemma with the Perron-Frobenius Theorem, it follows that

Lemma 4.1 (Finite Probabilistic Solution Lemma) *Let PE be a probabilistically tagged flat finite equation system (defined over ZFCU-+AFA). Then, PE has a unique solution.*

which has the immediate consequence that

Theorem 4.1 *Let PE be a probabilistically tagged flat finite equation system involving m indeterminates and k urelements (defined over ZFCU-+AFA). Then, one may construct a mean-consistent, mean-indexed infinite-order probability distribution S with $n = m + k$ elements F_i, one element corresponding to each indeterminate or urelement in PE, via defining $F_i(F_j) = w_{ij}$ where the latter is the (i, j) entry of the stochastic matrix W associated with PE (and setting the mean μ_i of F_i equal to the i'th entry of the stationary vector of W .*

In essence, what we have shown is that if one takes any finite system of set equations, and tags the constants and variables on the right hand sides of the equations in a reasonable way, one obtains a correspondent infinite-order probability distribution. To make the above arguments work for infinite systems of equations also seems feasible, but would require hassling with measures more complex than the counting measure and hence lengthen the argument, so we defer this to a future paper.

5 Uncertain Reasoning about Self-Referential Statements

The most obvious implications of this new sort of probability lie in the domain of *semantics*. Infinite-order pdf's allow us to intermix uncertainty with self-reference and mutual inter-reference in a manner that has not been possible using previous formalisms. This opens the door for various sorts of syntheses between semantic theories based on non-foundational sets (such as situation semantics, see [3]) and semantic theories founded on notions of uncertainty (e.g. [2]); though the exploration of such syntheses in detail would go beyond the scope of this brief paper.

As a simple example of knowledge that is most naturally represented using type-free relationships, consider the set of relationships:

$$N_1 = BensBeliefs$$

$$R_1 = love(Ben, BensBeliefs)$$

$$R_2 = believe(Ben, R_1)$$

$$R_2 \in N_1$$

Intuitively, what this expresses is: Ben loves his beliefs; Ben believes that he loves his beliefs; and finally, "Ben believes that he loves his beliefs" is one of Ben's beliefs. If we take the standard set-theoretic approach of defining relationships as ordered tuples, we then have $N1 < R1 < R2 < N1$, where $A < B$ denotes "A is a member of B, or a member of a member of B, etc." The axiom of foundation is violated. But introducing AFA and using the standard reduction of relations to sets, the above set of relationships can be reduced to a system of set equations and then to a flat finite equation system as defined above. One can then look at assignments of probability values to the terms in the relationship-system, such as $BensBeliefs$ and $believe$.

Finally, there is also an apparent connection to computational probabilistic inference. In [6] the author has developed a system for inference using second-order probabilities, which however is immediately seen to work identically on mean-consistent mipdf's as on second-order probability distributions. Type-free and typed relationships may be probabilistically reasoned about in a seamlessly interoperative way. We believe this may serve as a quite robust approach to practical reasoning about beliefs and social situations, but the details are beyond the scope of this paper.

References

[1] P. Walley, *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, 1990.

[2] J. Halpern, *Reasoning about Uncertainty.* MIT Press, 2003.

[3] J. Barwise and J. Etchementy, *The Liar.* Oxford University Press, 1989.

[4] P. Aczel, *Non-Well-Founded Sets*, CLSI Lecture Notes No. 14, Stanford University

[5] H. Minc, Nonnegative matrices, John Wiley and Sons, New York, 1988

[6] M. Ikle' and B. Goertzel. Indefinite Probabilities for General Intelligence, in *Advances in Artificial General Intelligence*, Edited by Ben Goertzel and Pei Wang. IOS Press.