Inheritance of Statistical Properties

Neil C. Rowe

Department of Computer Science
Stanford University
Stanford, CA 94305

Abstract

Statistical aggregate properties (e.g. mean, maximum, mode) have not previously been thought to "inherit" between sets. But they do in a weak sense, and a collection of such "weak" information can be combined in a rule-based architecture to get stronger information.

1. Motivation

Suppose we have conducted a census of all elephants in the world and we can definitely say that all elephants are gray. Then by set-to-subset inheritance of the "color" property, the set of elephants in Clyde's herd must be gray, Clyde's herd being some particular herd of elephants.

This will not work for statistical aggregate properties such as maximum and mean. Suppose our census found that the longest elephant in the world is 27 feet long, and the average elephant 15 feet. This does *not* mean the longest elephant in Clyde's herd is 27 feet, nor the average in the herd 15 feet. But a weak form of inheritance is present, for we can assign different degrees of likelihood to the following:

This work is part of the Knowledge Base Management Systems Project, under contract #N00039-82-G-0250 from the Defense Advanced Research Projects Agency of the United States Department of Defense. The views and conclusions contained in this document are those of the author and should not be interpreted as representative of the official policies of DARPA or the US Government.

1. "The longest elephant in Clyde's herd is 30 feet long."
2. "The average elephant in Clyde's herd is 30 feet long."
3. "The longest elephant in Clyde's herd is 27 feet long."
4. "The average elephant in Clyde's herd is 27 feet long."
5. "The longest elephant in Clyde's herd is 16 feet long."
6. "The average elephant in Clyde's herd is 16 feet long."

Statements 1 and 2 are impossible. Statement 3 is possible but a bit unlikely, whereas statement 4 is almost certainly impossible. Statement 5 is surprising and hence apparently unlikely, whereas 6 is quite reasonable. Since we don't know anything of Clyde's herd other than that they are elephants, a kind of inheritance from the properties of elephants in general must be happening.

The issue here is more important than elephants. Thousands of databases in existence support statistical questions about their contents. Exact answers to such questions may be very time-consuming for large data sets and/or remote access. Many users, especially non-statisticians, may be willing instead to accept much faster approximate answers via inheritance methods [5].

2. Our four-characteristic approach

We wish to address inheritance of the set properties maximum, mean, standard deviation, median, mode, fits to simple distributions, and correlations between different values of the same item. Our theory concerns set representation only (but sets of cardinality one can represent individuals). It concerns "definitional" sets primarily (those with absolute criteria for membership) as opposed to "natural kind" sets [1] (though degrees of set membership as in fuzzy set theory could be introduced). The theory mainly deals with extensions (exemplars), not intensions (meanings). It also only addresses the set-subset semantic relationship; however, often other relationships can be viewed this way by "atomization" of the included concepts, e.g. geographical containment may be seen as a set-subset relationship between sets of points.

The key is to note that while in a few cases statistical properties inherit values exactly from set to set, in most cases they do not; but that there are characterizations of a numeric statistic that will inherit much more often:

- an upper bound on its value
- a lower bound on its value
- a best estimate of the value
- a standard deviation of possibilities for the value

Some examples:

- An upper bound on the mean of a subset is the maximum of the set.
- A lower bound on the maximum of a subset is the minimum of the set.
- A best estimate of the mean of a subset, in the absence of further information, is the mean of the set.
- A standard deviation of the mean of a subset is approximately the standard deviation of the set times the square root of the difference of the reciprocals of the subset size and set size.

The last also illustrates an important feature of statistical property inheritance, namely that functions (in the mathematical sense) of values may be inherited rather than the values themselves. But since the different values are so strongly coupled it seems fair to still call it "inheritance".

Inheritance of nonnumeric statistics such as mode can analogously be characterized by a best estimate, a superset guaranteed to include all values, and an estimated relative frequency of the estimate among all possible values. Note this approach is a distinct alternative to often-arguable certainty factors for specifying partial knowledge.

3. Inheritance types

There are three "dimensions" of statistical inheritance: what statistic it concerns, which of the four abovementioned manifestations it addresses, and how it basically works. The main categories of the latter are:

- Downwards inheritance. That is, from set to subset, as in the examples of the last section. This is the usual direction for statistical inheritance since it is usually the direction of greatest fanout: people tend to store information more for general concepts than specific concepts, for broadest utility. In particular, downwards inheritance from sets to their intersection is very common in human reasoning, much more so than reasoning with unions and complements of sets.
- Upwards inheritance. Inheritance from subset to set occurs with set unions, in particular unions of disjoint sets which (a) seem easier for humans to grasp, and (b) have many nice inheritance properties (e.g. the largest elephant is the larger of the largest male and largest female elephants). Sampling, random or otherwise, to estimate characteristics of a population is another form of upwards inheritance, though with the special disadvantage of involving a non-definitional set. Upwards inheritance also arises with caching [4]. People may cache data on some small subsets important to them (like Clyde's herd) in addition to general-purpose data. Upwards (as well as

downwards) inheritance is helpful for dealing with "intermediate" concepts above the cache but below general-purpose knowledge (e.g. the set of elephants on Clyde's rangelands).

- Lateral inheritance. A set can suggest characteristics of sibling sets of the same parent superset [2]. Two examples are set complements (i.e. the set of all items not in a set, with respect to some universe), and when sibling sets differ only by an independent variable such as time or space, and there are constraints on the rate of change (i.e. derivatives) of numeric attributes between siblings (e.g. the stock market average on successive days).

- Diagonal inheritance. An interesting hybrid of downwards and lateral inheritance is possible with statistical properties. Given statistics on the parent and some set of siblings, we can often "subtract" out the effect of the known siblings from the parent to get better estimates on the unknown siblings. For instance, the number of female elephants is the total number of elephants minus the number of male elephants. This also works for moment and extrema statistics.

- Intra-concept inheritance. Inheritance can also occur between different statistics on the same set, if certain statistics are more "basic" than others. For instance, mean can be estimated as the average of maximum and minimum, and thus can be said to "inherit" from them; people may reason this way, as in guessing of the center of a visual object from its contours. But in principle almost any direction is possible with numerical and nonnumerical relaxation techniques.

- Value-description-level inheritance. Real-world property values, especially nonnumeric ones, can be grouped at different levels of detail, and inheritance is possible between levels for the same set and same statistic. For instance, the number of different herds can be estimated from the number of different elephants and general knowledge of how many elephants are in a herd.

- Inheritance-rule inheritance. Some sets are sufficiently "special" to have additional inheritance rules for all subsets or supersets. An example is an all-integer set, where for any subset an upper bound on the number of distinct values for that property is the ceiling on the range.

4. Closed-world inferences

Since there are many statistics, and even a small set can have many subsets, default reasoning is essential for efficiency with statistical properties. Inferences from the absence of explicit memory information are common in human reasoning [3]; in particular, the idea that "sufficiently important" sets whose statistics are not explicitly noted must be not "unusual" in regard to those statistics. We can define "sufficiently important" and "unusual" relative to what inheritance predicts.

5. A production system architecture

So many different kinds of inheritance (even just those applicable to the same concept), complicated combination and cascading of different inheritances, inheritance of functions of values rather than values, inheritance- inheritance -- all this classically suggests a production system architecture is needed. That is, the encoding of inheritance categories as production rules.

There are two conflict resolution issues for the control structure of such an architecture: which rules to invoke, and how to resolve different answers from different rules.

Many different inference paths can be followed in making a statistical estimate, even not including all the possible rearrangements of a set expression involving intersections, unions, and complements. Since these can give different final answers, it's important to explore as many of these in parallel as possible, unlike most production systems where a single "best" alternative is desired. But some limits to parallelism have to be set for complicated queries, and we are currently investigating "weakest-first" inference. (Arithmetic must be generalized for operations on intervals.)

Combining results from different inference paths is straightforward for numeric statistics. Intersect the ranges to get a cumulative range. Get the cumulative estimate by assuming independence for all estimates, combining as if their errors were characterized by normal distributions via the classical statistical formulas; and the cumulative standard deviation follows directly. Even with nonindependence in the latter calculations the estimate should not be off much, and the standard deviation for the two-path case is never more than 70% ($2^{1/2}$) of what it should be.

6. An application

We are implementing a program that uses these ideas to answer statistical questions for a large database [5]. It uses several hundred rules from a variety of sources: mathematical definitions, extreme-value analysis of definitions, statistical theorems, exploratory data analysis, database dependency theory, statistical database inference security research, psychology of conceptual classes, and general principles of information systems. As with many other "expert systems" in artificial intelligence, there is more fundamental mathematical theory -- in this case, nonlinear optimization and cross-entropy minimization [6] -- that underlies many of the rules, but is too intractable for all but the simplest cases to be of much use.

References

1. R. J. Brachman and D. J. Israel. KL-ONE Overview and Philosophy. In *Research in Knowledge Represenation for Natural Language Understanding: Report No. 4785*, W. A. Woods, Ed.,Bolt Beranek and Newman, 1981, pp. 5-26.

2. Jaime G. Carbonell. Default Reasoning and Inheritance Mechanisms on Type Hierarchies. Proceedings, Workshop on Data Abstraction, Databases, and Conceptual Modelling, Pingree Park CO, June, 1980, pp. 107-109.

3. Allan Collins. Fragments of a Theory of Human Plausible Reasoning. Proceedings, Second Conference on Theoretical Issues in Natural Language Processing, Urbana IL, July, 1978, pp. 194-201.

4. D. B. Lenat, F. Hayes-Roth, and P. Klahr. Cognitive Economy. Working Paper HPP-79-15, Stanford University Heuristic Programming Project, June, 1979.

5. Neil C. Rowe. Rule-Based Statistical Calculations on a Database Abstract. Proceedings, First LBL Workshop on Statistical Database Management, Menlo Park CA, December, 1981, pp. 163-176.

6. John E. Shore and Rodney W. Johnson. "Properties of Cross-Entropy Minimization." *IEEE Transactions on Information Theory IT-27*, 4 (July 1981), 472-482.