Statistical Inference as Default Reasoning

Henry E. Kyburg, Jr.
kyburg@cs.rochester.edu
Computer Science and Philosophy
University of Rochester
Rochester, NY 14627, USA

Choh Man Teng
teng@cs.rochester.edu
Computer Science
University of Rochester
Rochester, NY 14627, USA

Abstract

Classical statistical inference is nonmonotonic in nature. We show how it can be formalized in the default logic framework. The structure of statistical inference is the same as that represented by default rules. In particular, the prerequisite corresponds to the sample statistics, the justifications require that we do not have any reason to believe that the sample is misleading, and the consequent corresponds to the conclusion sanctioned by the statistical test.

Classical Statistical Inference

Classical statistical inference, which is to be contrasted with Bayesian statistics, is a pervasive form of uncertain inference. Most statistical inference is done in a classical framework. It *is* inference, despite what some authorities say.[1] Furthermore, it is nonmonotonic. New data — an expansion of the premises — could lead to the withdrawal of the conclusion. In fact it is this very nonmonotonicity that leads some writers to deny that nonmonotonic logic is logic at all (Morgan 1997).

As the practitioners of classical statistical inference insistently remind us, it is NOT probabilistic: the outcome of a statistical inference is not the assignment of a probability to a statistical hypothesis, but the categorical acceptance or rejection of such an hypothesis.[2] To be sure, inferences of this form are often characterized by a number: a significance level, a confidence coefficient, a size,... But while this may make it reasonable to assimilate such inferences to the class of "quantitative" inference, that is not at all the same as to force them into the mold of assigning probabilities to statements. This would be resisted by the majority of statisticians — all but the "Bayesians", who are still a minority. On the other hand, these inferences fit rather neatly into the framework of nonmonotonic logic. The validity of such inferences strongly depends on the fact that certain items of information are *not* in our body of knowledge.

[1]Neyman (Neyman 1950), Lehman (Lehman 1959), and others insist that it is not inference, but decision. But even those authors must admit that the "decision" that is the result of research is often a decision to believe, or to publish rather than to take some more direct action.

[2]Here is a sampling of texts in which this classical viewpoint is affirmed: (Neyman 1950; Fisher 1956; Mood & Graybill 1963; Lehman 1959; Alexander 1961; Wilks 1962).

Example

The robot fireman needs to keep track of the ambient temperature of the air that surrounds him. He is equipped with a temperature measuring device that has been calibrated. The corrected temperature reading is T_r. Our firerobot may infer that the true ambient temperature is in the interval $T_r \pm 1.96\sigma_r$, where σ_r is the standard deviation characteristic of measurements made with the instrument he employs. Why is this a reasonable inference for him to make? Why should we design our firerobot to make this inference?

As is the case with most measurements, it may be assumed that the distribution of errors of measurements is normal, with a mean of 0 (for corrected readings) and a variance ${\sigma_r}^2$. There are many reasons for this even beyond the simplicity of that assumption; ever since Gauss there have been arguments that purport to show that minimal assumptions about the measurement process lead to this conclusion. In most cases the measuring instrument can be calibrated so that the mean error of a "corrected" measurement can be taken to be 0. The mean of the population of possible measurements is the *true value* of the quantity being measured. Therefore that true value, the mean, is just what we want to obtain confidence limits for.

R. A. Fisher, one of the founding fathers of modern statistics, held that the inference to the mean of a normal distribution was one of the few cases in statistics in which "inverse inference" could be made to work. The crucial factor is that there exists a "pivotal quantity" with known distribution independent of the value of the unknown quantity. Suppose that the quantity X is normally distributed with known standard deviation σ_X and unknown mean μ in the population P. It follows that the quantity $|X - \mu|$ is normally distributed with mean 0 and standard deviation σ_X in that same population. Let us now take a random sample consisting of a single member a of the population P, and observe the value of $X(a)$.

According to Fisher, *provided that a is not known to belong to any specifiable subset of the population about which we have different information*, the existence of a pivotal quantity entitles us to derive a "fiducial" probability density for μ that is normal, has mean $X(a)$, and variance σ_X^2. From this density we can calculate the interval of minimum length that has a *fiducial* probability of 0.95 of including the value of μ: $X(a) \pm 1.96\sigma_X$.

Many statisticians find Fisher's notion of "fiducial probability" unclear; most insist that "0.95" does not represent a probability at all, but a "confidence": our confidence that in the long run the procedure outlined will not, in a particular case, lead us astray (Walpole & Myers 1978, 195) (Mood & Graybill 1963, 249). Thus it is the *procedure* of forming a 0.95 confidence interval, under the circumstances outlined, that is characterized by a long run frequency (a genuine frequency probability) of success of 0.95. More precisely, a 95% confidence interval for a parameter is an interval-valued function of the sample that will cover the parameter 95% of the time in the long run. This is a property of the *function*, not of its value for a particular sample.

Mood writes "Confidence intervals and regions provide good illustrations of uncertain inferences." (Mood & Graybill 1963, 151) This is an understatement. Almost every quantitative statement we make is based on one or more measurements. Every measurement is subject to error. In almost all cases, errors of measurement are assumed to be distributed normally with a known mean (usually, but not always, taken to be 0) and a variance σ^2 taken to be a known characteristic of the method of measurement. In such cases, the true value of a quantity being measured is the mean of a normally distributed quantity (measurement observations) having a mean equal to that true value, and a variance equal to that same σ^2. The mean is the true value, and what we can (uncertainly) infer is that it falls in a certain interval: a confidence interval determined by our observation and the level of confidence we want to employ.

Thus whenever the result of measurement enters into our deliberations in planning or acting or deciding (or anything else in AI) it does so as an instance of confidence interval estimation of the mean of a normally distributed quantity. Whether we refer to the level as a "confidence" or a "probability" or a "fiducial probability" the import of the inference is clear: infer that the quantity in question lies in the interval calculated, pending further evidence.

But how are we to understand the italicized proviso? Statisticians have not been clear, and have often left it as a matter of common sense or good statistical practise. We will argue that it functions precisely as a justification in a default rule, and that doing so allows us to construe the inference as defeasible and to make its conditions of defeat clear.

Default Logic as a Framework for Statistical Inference

Default logic provides a formalism within which the principles of statistical inference can be codified and evaluated. It provides a general framework within which particular applications of statistical inference, for example measurement, can be handled as nonmonotonic inferences. We can spell out the conditions under which an observation supports a claim about the value of a quantity, and the conditions under which that claim must be withdrawn, despite the observation.

Default Logic

Default logic is intuitively appealing and easy to understand, at least at first sight. There are some subtleties that need to be addressed, but for the time being, let us leave them aside. We refer to "default logic" in a very loose way, as a prototypical non-monotonic framework which allows us to express various reasoning methods in terms of the uniform syntax of default rules. We do not necessarily follow the semantics of the specific version of default logic as advocated by Reiter (Reiter 1980), or for that matter any particular variant of default logic proposed in the literature (Lukaszewicz 1988; Brewka 1991; Gelfond *et al.* 199 1; Delgrande, Schaub, & Jackson 1994; Mikitiuk & Truszczyński 1995, for example).

A *default rule* d is an expression of the form $\frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma}$, where α, $\beta_1, \ldots, \beta_n$, γ are logical formulas. We call α the *prerequisite*, $\beta_1, \ldots, \beta_n$ the *justifications*, and γ the *consequent* of d. A *default theory* Δ is an ordered pair $\langle D, F\rangle$, where D is a set of default rules and F is a set of logical formulas.

Loosely speaking, a default rule $\frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma}$ conveys the idea that if α is provable, and $\neg\beta_1, \ldots, \neg\beta_n$ are not provable, then we by default assert that γ is true. For a default theory $\Delta = \langle D, F\rangle$, the known facts constitute F, and a theory extended from F by applying the default rules in D is known as an *extension* of Δ. Basically, a default extension contains the set of given facts, is deductively closed, and all default rules that can be applied in the extension have been applied. In addition, an extension has to be minimal, that is, every formula in an extension either is a fact or a consequent of an applied default rule, or is a deductive consequence of some combination of the two.

The Prerequisite and Justifications

A default rule, $\frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma}$, can be applied to conclude the consequent γ when the conditions associated with its prerequisite α and justifications $\beta_1, \ldots, \beta_n$ are satisfied. The prerequisite condition α is satisfied by showing that α is "present", and each of the justification conditions β is satisfied by showing that $\neg\beta$ is "absent".

In the classical logic framework (which is what default logic is based on), the presence or absence of a formula is determined by deductive provability: α

is "present" iff α is provable from a set of sentences, and $\neg\beta$ is "absent" iff $\neg\beta$ is not provable from the same set of sentences. However, logical provability need not be the only way to determine whether a formula is "present" or "absent." In particular, formulas obtained by the application of default rules may qualify as "present."

This is particularly important in the present context, as we will see when we examine the justifications for statistical inference.

Justifications vs. Assumptions

One advantage of the default approach to statistical inference can be brought out by focusing on the distinction between a "justification" and an "assumption". An *assumption* is a statement that we add to our body of knowledge, and use as if it were part of our data or set of premises. The justifications $\beta_1, \ldots \beta_n$ of a default need not be treated as *positive* knowledge: their role is hypothetical: *If* we do *not* know that any of them are *false*, then we can go ahead with our inference. The acceptability of a *default* depends on the fact that we do *not* know something, rather than on the fact that we are pretending to know something. In a default rule, the justification β need never occur in our body of knowledge, or as a premise. We require only that its *negation* not be either deductively or nonmonotonically acceptable.

For example, in making an inference from a measurement to the true value of a quantity, we are making an inference from a sample of size one to the mean of a normal distribution. We *assume* that the errors of measurement are distributed approximately normally. But we need as *justification* the fact that we do *not* know that there is anything wrong with that measurement: that the balance hasn't been dropped, that the measurement wasn't made by a notoriously sloppy worker, etc. The assumption of approximate normality is a (presumably) well justified premise. That there is nothing wrong with the measurement in question is not a premise, but a *justification* in the default sense. We do not have to know that the measurement was made by a careful worker; it is only necessary that we do *not* know that it was made by a sloppy worker.

This has been somewhat confused by our tendency to focus on "normal" defaults, in which the justification is the same as the conclusion, for example,

$$\frac{\text{Tweety is a bird : Tweety flies}}{\text{Tweety flies}}.$$

But there are many well known defaults that are not normal, for example[3],

$$\frac{\text{Tweety is an adult : Tweety is not a student}}{\text{Tweety has a job}}.$$

We do not, at any point, add "Tweety is not a student" to our premises.

[3](Reiter & Criscuolo 1981, a variation of).

Assumptions are sometimes invoked in statistical inference. An example is the "Simple Random Sampling" (Moore 1979) assumption often mentioned, and construed as the claim that each equinumerous subset of a population has the same probability of being drawn. According to this assumption, every sample must have the same chance of being chosen; but we *know* that that is false: samples remote in space or time have no chance of being selected.

We cannot choose a sample of trout by a method that will with equal probability select every subset of the set of all trout, here or there, past or present, with equal frequency. Yet the *population* whose parameter we wish to evaluate may be precisely the set of all trout, here and there, past and present.

This is also true of the sampling assumption mentioned by Cramér (Cramér 1951, 324) (and also by Baird (Baird 1992, 31)) which requires that each element in the domain has an equal probability of being selected.[4] What is really required (perhaps among other things) is that we *not know* that there is something special about the sample that vitiates the conclusion we hope to draw from it. This is the standard form of a "justification" in default logic, which requires that we do *not* know something.

We cannot, therefore, take simple random sampling as a *premise* or an *assumption* of our statistical argument. We not only have no reason to accept it, but, usually, excellent reasons for denying it. But the arguments go through anyway, and justifiably so. The structure is the classical default structure: Given the prerequisite of a sample statistic, we may infer the conclusion with confidence $1-\alpha$, *provided* that it is possible for all we know that ... what? Not that the sample is a simple random sample, for we know it is not. Not that the population parameter is in the interval $m/n \pm \epsilon$, because this would be possible even if we knew the sampling method to be biased. Not that the sample *might* have been selected by a simple random sampling method, for this is true of all samples.

We might ask that it must be possible, relative to what we know, that the sample is a *good* sample, or an unbiased sample, or a *mere member* of the class of equinumerous samples. But this is not quite right. A grossly biased sampling method *could* yield just the same sample as a perfectly random method (if such were possible). What is required is that we have no *reason to believe* that the sampling procedure is biased, where the procedure is biased just in case it has long run properties that undermine the applicability of the statistics on which we are basing our inference.

More specifically, it must be possible for all we know that the sample belongs to no subclass of the set of equinumerous samples that would serve as a reference class for a *conflicting* inference regarding representativeness, that is, we have *no reason to believe* that the

[4]Actually this condition is not sufficient; we also must require independence of selections.

sample is drawn from a subclass in which the frequency of representative samples is less than that among the set of samples on which the statistical inference is based (ordinarily the set of all equinumerous subsets of the population). If we draw a sample of coffee beans from the top of a bin in order to make an inference concerning the proportion of bean fragments, we will almost surely be wrong: the cracked beans will fall to the bottom on the drawer.

While we should take precautions in sampling, we should do so, not because we can *ensure* getting a good sample, but because we can in that way defend ourselves against certain kinds of bad sample. The mere possibility of having gotten a bad sample should not inhibit our inference. In the absence of *evidence* to the contrary, the inference goes through. When there is evidence against fairness -- and note that this may be *statistical* evidence — the inference is blocked.

The Example Worked Out

Consider the inference to the mean of a normal distribution when we regard the standard deviation as known; for example the case of the robot fireman. Suppose we want to conclude with confidence 0.95 that Q — the ambient temperature — lies in the interval $V \pm 1.96\sigma$. The prerequisite is that our observed value is V. The justifications might be that:

- This is the only relevant data we have concerning Q; otherwise we should also take account of that other data.
- We have no reason to believe that the measuring instrument is not well calibrated.
- We have no reason to think that the observation was careless.
- We have no reason to think that the observation was made in an abnormally warm or abnormally cool part of the space we want to characterize.
- We have no reason to think that the sample was atypical in any other way.

Each of these defaults could be made the basis for a defeating default: that is, for example, if we have other data concerning Q, then we should take that other information into account, and not base our inference only on the single observation.

The second justification might be thought of as a justified assumption, but often we do not calibrate our instruments: we take them to be well calibrated unless we have reason to believe otherwise.

Third, if we know that the measurement was made by a notoriously sloppy technician, we will not use it as a basis for believing that $Q \in V \pm 1.96\sigma$.

Fourth, if we are seeking the average temperature in a closed vessel, we will avoid measurements near the walls of the vessel, or near heating elements, or ....

Fifth, if the result of the observation was an observation recorded as 312 degrees Centigrade, but we know that liquid water was present and the pressure was one atmosphere, we will not perform the suggested inference: we will know that something is fishy.

The nonmonotonic rule of inference might be expressed in some such form as this:

$$\frac{X(a) = x \wedge |\mu - X| \text{ is } N(0, \sigma^2) : \beta_1, \ldots, \beta_n}{X(a) - 1.96\sigma \leq \mu \leq X(a) + 1.96\sigma},$$

where the β's represent justifications of the kinds just suggested. Thus, given a default theory, $\Delta = \langle D, F \rangle$, where D contains the above default rule, and F contains, among other information, $X(a) = T_r$, $|Q - X|$ is $N(0, \sigma_r^2)$, we may infer with confidence 0.95 that $Q \in T_r \pm 1.96\sigma_r$, provided that none of the justifications are found to be false in the default extension.

Note that these justifications need not take the form of deductive consequences from our database. What is required is that we not ***have reason to believe*** the negations of any of them. The evidence that gives us reason may in fact have the same source as the evidence we would base our default inference on. For example we might find in a large sample from a population that there is *internal* evidence that the sample is biased in a relevant way.

Conclusion

Characterizing statistical inference and other sorts of specialized reasoning methods in the default logic framework is beneficial in several ways. The framework has a well defined syntax which provides a versatile and uniform platform for incorporating diverse reasoning techniques. A reasoning process is broken down into steps formalized by default rules. The prerequisite of the default rule represents the pre-conditions that need to be true of the reasoning step; the justifications represent the conditions that underlie the validity of the reasoning step. These justifications are typically implicit and given little attention to in normal practice. However, the justification slot in the default rule highlights their importance and make them explicitly available for examination. The default logic structure also makes it easier to keep track of all the justifications accumulated during the course of a reasoning process. The set of collective justifications is readily accessible for scrutiny when the reasoning process or the results seem questionable.

The specialized reasoning methods of particular interest are the non-logical "foreign" methods which have been developed tailored to some particular classes of problems. Classical statistical inference is an example of such a specialized quantitative method, designed to draw conclusions about some population characteristics based on samples drawn from this population. These non-logical methods and their specialty areas are normally not considered in the study of logical systems. However, by casting these methods as default rules, we can assimilate them into the logical framework. Having a collection of specialized tools vastly expands the domain of the reasoner, or at least allows the reasoning tasks to be done more effectively by using the most appropriate technique, and in a more integrated way by having a uniform structure of all the techniques.

We can think of the classical logic machinery that default logic is based on as the meta-problem solver. We can make use of all the techniques developed for classical logic for general problem solving. What classical logic cannot solve easily or quickly, we can invoke specialized default rules derived from extra-logical methods to achieve the desired result. This two-tier arrangement can be built in a modular way, in the sense that new reasoning methods can be transformed into default rules and plugged into the meta-system with minimal restructuring.

One may argue we already have lots of fine statistical programs, which can calculate whatever statistics we fancy. So what are we doing here?

It is true that the statistical programs that are widely available nowadays can give us statistical results efficiently. However, performing the calculations is only the more trivial part of the task of statistical inference. The programs can do their job *only* when we have decided what to test, how to test it, and, after the calculations, how to interpret the results. In other words, we need to supply the program with precise instructions on what tests to administer and what statistics to collect. The statistical program itself is not concerned about whether the test chosen is appropriate; it just carries out the calculations.

What we are dealing with here are exactly those parts in the inference process that are outside of the scope of a statistical program. Statistical programs should best be thought of as "statistical assistants", one of the tools an autonomous agent has for reasoning about the world. Formalizing the statistical inference process in terms of default rules provides a way for specifying the conditions under which an agent can perform a statistical inference, in the same way that deductive inferences are extracted from its "theorem prover assistant". The statistical assistant serves the role of connecting the information provided in the prerequisite of the inference default to the conclusion of the default. This leaves entirely open the applicability of the default. The justifications of the default determine whether or not the agent should accept the conclusion of the inference. Thus we are not proposing a competing statistical program; we are formalizing the rules of inference that can make use of these programs.

Acknowledgement

Support provided by NSF grant IRI-9411267 is gratefully acknowledged.

References

Alexander, H. W. 1961. *Elements of Mathematical Statistics.* New York: John Wiley.

Baird, D., ed. 1992. *Inductive Logic: Probability and Statistics.* Englewood Cliffs, N.J.: Prentice-Hall.

Brewka, G. 1991. Cumulative default logic: In defense of nonmonotonic inference rules. *Artificial Intelligence* 50:183–205.

Cramér, H. 1951. *Mathematical Methods of Statistics.* Princeton.

Delgrande, J. P.; Schaub, T.; and Jackson, W. K. 1994. Alternative approaches to default logic. *Artificial Intelligence* 167–237.

Fisher, R. A. 1956. *Statistical Methods and Scientific Inference.* New York: Hafner Publishing Co.

Gelfond, M.; Lifschitz, V.; Przymusinska, H.; and Truszczyński, M. 1991. Disjunctive defaults. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, 230–238.

Lehman, E. 1959. *Testing Statistical Hypotheses.* New York: John Wiley and Sons.

Łukaszewicz, W. 1988. Considerations on default logic: An alternative approach. *Computational Intelligence* 4(1):1–16.

Mikitiuk, A., and Truszczyński, M. 1995. Constrained and rational default logics. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.*

Mood, A. M., and Graybill, F. A. 1963. *Introduction to the Theory of Statistics.* New York: McGraw-Hill.

Moore, D. S. 1979. *Statistics.* San Francisco: W. H. Freeman and Co.

Morgan, C. 1997. Non-monotonic logic is impossible.

Neyman, J. 1950. *First Course In Probability and Statistics.* New York: Henry Holt And Co.

Reiter, R., and Criscuolo, G. 1981. On interacting defaults. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, 270–276.

Reiter, R. 1980. A logic for default reasoning. *Artificial Intelligence* 13:81–132.

Walpole, R. E., and Myers, R. H. 1978. *Probability and Statistics for Engineers and Scientists, 2'nd Edition.* New York: Macmillan.

Wilks, S. 1962. *Mathematical Statistics,.* New York: Wiley.