Circling the Truth: Model Selection Criteria as a Metric of Verisimilitude in Theory Selection

By K. Raleigh Hansen (Washington State University)

Abstract: The purpose of this research is to investigate the possibility of using aspects of model selection theory to overcome both a logical problem and an epistemic problem that prevents progress towards the truth to be measured while maintaining a realist approach to science. Karl Popper began such an investigation into the problem of progress in 1963 with an idea of verisimilitude, but his attempts failed to meet his own criteria, the logical and epistemic problems, for a metric of progress. Although philosophers have attempted to fix Popper's verisimilitude, none have seemed to overcome both criteria yet. My research analyzes the similarities between Predictive Accuracy (PA) and Akaike's Information Criterion (AIC), parts of model selection theory, and Popper's criteria for progress. I find that, in ideal data situations, it seems that PA and AIC satisfy both criteria; however, in non-ideal data situations, there are issues that appear. These issues present an interesting dilemma for scientific progress if it turns out our theories are in non-ideal data situations, yet PA and AIC seem to be better overall indicators of scientific progress towards the truth than other attempts at overcoming the problems of Popper's verisimilitude.

One problematic issue when discussing scientific progress is whether or not our current theories have made any progress towards the truth or have just become better predictive tools. There is an intuitive notion that newer theories are truer than older theories because they appear to identify more true causes of a target system. However, it turns out that it is notoriously difficult to provide an analysis of what it means for one theory to be closer to the truth than another theory. The issue is even more pronounced when considering the pessimistic induction: since all of our past theories have been false, it is likely that all of our current will also be false and perhaps our future theories as well. This poses a problem for scientific realism which holds that identifying the true causes of a target system is an important aim of science.

While the discovery of new causes that affect target systems do seem to be an important part of scientific progress, it is not clear that increasing the ability to predict the behavior of target systems must always account for more causes known to affect the target system (Forster and Sober 1994). In fact there is some evidence that our best predictive models and theories might not always be our best explanatory models and theories (Goldsby 2013). However, if we want to define progress in realist terms, there needs to be some account of what proximity to the truth is and how newer theories get us closer to the truth. I will refer to these two concerns as the logical problem and the epistemic problem respectively.

An early attempt to overcome the logical and epistemic problems was introduced by Karl Popper in his work *Conjectures and Refutations*. Popper (1963) called his attempt to overcome the two problems verisimilitude. The concept behind verisimilitude is intuitive in nature – a theory is closer to the truth if it makes more true claims and fewer false claims – but his later commentators would point out critical flaws such that verisimilitude is not able to solve either the logical or epistemic problem. A number of attempts have been made to revise or fix Popper's language to make verisimilitude work, but none have overcome both the logical and epistemic problems. However, if progress can be defined as overcoming the logical and epistemic problems, then it is possible there may exist a framework elsewhere that satisfies that criteria.

One possible framework, predictive accuracy (PA), is a measure of the ability of a model to predict new data given old data. One plausible assumption is that the true model will be maximally predictively accurate, so increasing predictive accuracy will get one closer to the truth. According to Forster and Sober (1994), PA may be estimated using Akaike's Information Criterion (AIC). If PA can be a measure of closeness to the truth, then using a model selection framework like AIC can select models closer to the truth. If, in turn, AIC can select a model that is closer to the truth and is more predictively accurate than competing models, AIC can be useful for estimating progress. In this way, PA overcomes the logical problem by explaining how one model can be closer to the truth than another, and AIC overcomes the epistemic problem by showing that, when a new model is selected, it is because of both its increased proximity to the truth as well as its ability to predict new data.

The main concern for this paper is to investigate whether PA and AIC actually can overcome the logical and epistemic problems. I will begin by explaining why a notion of verisimilitude is important for the progress of science. I will then provide some background to Popper's account verisimilitude, and I will introduce model selection theory and explain how PA and AIC appear to satisfy the criteria demanded by verisimilitude. I will argue that PA and AIC can overcome both problems while in an ideal data situation and discuss what may occur while in non-idea data situations. Finally, I will address the problems of PA and AIC as a form of verisimilitude and discuss what sort of progress we may actually have made.

Why is Progress Towards the Truth Important?

There are two basic accounts of the goals of scientific inquiry: realism and instrumentalism. Scientific realism maintains a concern for understanding the truth behind phenomena including things that can't be directly observed. Even if the pessimistic induction is right, realism holds that newer theories can be closer to the truth than older theories. For example, it seems correct to say that even though Copernicus's heliocentric model of the solar system is false, it is still closer to the truth than Ptolemy's geocentric model.

Unlike realists, instrumentalists view scientific theories as tools that help capture or predict observable phenomena regardless of the truth-value of the theories themselves (Chakravartty 2014). In this way, an instrumentalist values theories that can predict or account for observable phenomena even if we can't know the truth about the unobservable commitments of that theory (Van Fraassen 1980). Instrumentalists believe that the truth of unobservables is inaccessible and science should be aimed predicting observable phenomena rather than identifying all and only true causes.

Although Popper was a realist, his critics would point out that his hypothetico-deductive approach[243] to science by falsifying theories only winnows away at a possibly infinite set of theories and doesn't allow for actual progress. Popper's (1963) verisimilitude was his attempt to show that false theories could have degrees of closeness to the truth, and that removing false theories does constitute progress towards the truth. Popper hoped that verisimilitude would allow him to be a realist while still holding to his hypothetico-deductive approach to scientific inquiry. If progress towards the truth is the goal of science as Popper claims, then discarding an instrumentalist approach is an important step.

Popper's Verisimilitude

Popper correctly identified the logical and epistemic problems that must be overcome for verisimilitude to provide a measure of progress. The aims of verisimilitude can be easily formulated as the following questions:

> (A) Can we explain how one theory can be closer to the truth, or has greater verisimilitude than another?
>
> (B) Can we show that scientific change has sometimes led to theories which are closer to the truth than their predecessors? (Forster; ms) [244]

The first question addresses the logical problem such that we must have an account of when one theory is closer to the truth than another. The second question addresses the epistemic problem. Given our epistemic limitations, we must be able to determine that the selection of one theory over another is actually progress towards the truth.

Of course, Popper had to clarify how the degrees of truth would be measured. Popper's (1963) intuitive definition of verisimilitude, *Vs*, of theory *A* is based upon a measure of the true and false contents

[243] Popper's (1959) hypothetico-deductive approach was presented in his *Logic of Scientific Discovery*. According to Popper's method, a hypothesis should be formed in a way that can be deductively falsified rather than supported by evidence.
[244] Forster credits an unpublished manuscript by Alan Musgrave for this formulation of the logical and epistemic problems.

of *A*. The Ct(*A*) is made of all of the logical consequences of *A* and can be divided into truth content, $Ct_T(A)$, and false content, $Ct_F(A)$. Truth content of *A* is the set of all claims that are true in Ct(*A*), and false content is the set of all claims that are false in Ct(*A*). $Ct_F(A)$, subtracted from $Ct_T(A)$ provides a measure of verisimilitude:

$Vs(A) = Ct_T(A) - Ct_F(A)$ (Popper 1963, 234)

This intuitive definition provides the basic notion behind verisimilitude within a single theory by discovering the number of true and false logical consequences of theory *A*. The intuitive notion behind this measure is simple; it provides a measure Vs(*A*) based upon $Ct_T(A)$ and $Ct_F(A)$. By quantifying the true and false content of theories, this definition would allow two theories, *A* and *B*, to be compared as follows:

$$Vs(A) > Vs(B) \leftrightarrow [Ct_T(A) - Ct_F(A)] > [Ct_T(B) - Ct_F(B)]$$

The intuitive definition is a good first pass at the logical problem, but real theories are more complicated. For example, assume there are two theories, *A* and *B*, and that theory *A* and theory *B* are both false. To explain this concept, Popper (1963) offers the following example for any given theory: assume that today is Monday and theory *A* states that today is Tuesday; although theory *A* is false, it still entails true logical content such as today is not Wednesday and today is either Monday or Tuesday (Popper 1963). Because there are an infinite number of consequences, the Popper's first pass can't actually serve as a measure of verisimilitude.

Popper improved upon his first pass by using set-theoretic terms to create a contrastive definition of verisimilitude. Popper's (1963) contrastive verisimilitude (PCV) can be stated as follows:

(PCV) $Vs(A) < Vs(B) \leftrightarrow [Ct_T(A) \subset Ct_T(B)] \wedge [Ct_F(B) \subseteq Ct_F(A)]$

That is to say that for *B* to have greater verisimilitude, *B* must make every true claim made by *A* and at least one additional true claim not made by *A*, and every false claim made by *B* must also be made by *A* without any additional false claims.

As an example, consider Ptolemaic astronomy and Copernican astronomy. For the sake of simplicity, suppose that the only difference in content between Ptolemaic astronomy and Copernican astronomy is the location of the sun and the Earth. Copernican astronomy makes one true claim not made by Ptolemaic astronomy, the Earth revolves around the sun. Ptolemaic astronomy makes one false claim not made by Copernican astronomy, the sun revolves around the Earth. If PCV holds, Copernican astronomy has greater verisimilitude because it makes all the true claims that Ptolemaic astronomy makes plus an additional true claim, all the false claims made by Ptolemaic astronomy are also made by

Copernican astronomy, and Copernican astronomy makes one fewer false claim. Popper had examples like this in mind when he developed PCV to satisfy the criteria for verisimilitude.

The Problem with Popper's Verisimilitude

PCV, however, is also problematic in a similar manner to Popper's intuitive definition. Working independently, Pavel Tichý (1974) and David Miller (1974) both discovered a critical logical flaw to PCV. Tichý (1974) and Miller (1974) both pointed out that two competing false theories will never meet the subset relations PCV lays out because whenever a new true consequence is added, a new false consequence is added as well. Consider the following claims:

P1: The sun revolves around the earth

P2: The planets move in perfect circles

C3: The Earth revolves around the sun

Of course, we now know that P1 and P2 are false and C3 is true. The Ptolemaic model says P1 and P2 are true. The Copernican model says P2 and C3 are true. Now consider the following claim:

C4: P2 and C3

C4 is false because any conjunction that contains one false conjunct is always false. It is also a false claim that is not contained within the Ptolemaic theory. This can be called the conjunction problem. A true claim made by theory *B* but not made by theory *A* can be conjoined with a false claim made by theory *B* to create a new false claim not made by theory *A*. Thus, PCV will fail.

The incomparability of false theories is one of the consequences that developed from analysis of Popper's theory of verisimilitude. Tichý's (1974) and Miller's (1974) treatments of Popper's work show that it is impossible to add true consequences to a theory without also adding false ones, and equally impossible to subtract false consequences without also subtracting true ones. Two theories cannot be compared in terms of scientific progress towards the truth as Popper has defined it either as an intuitive notion or through PCV.

Applying Model Selection as Verisimilitude

If the concern of verisimilitude is to produce results that show theory progression is moving towards the truth by overcoming the logical and epistemic problems, it may be possible to look to forms of model selection that may serve the same purpose. A model is simply a set of equations that contain a number of adjustable parameters that is used to explain or predict a phenomenon (Forster 2000). A model can be broken down into the following parts: parameters, variables, and error terms. Consider the following toy models:

$$(M1) \quad y = ax_1 + e$$

$$(M2) \quad y = ax_1 + bx_2 + e$$

$$(FIT) \quad y = 7x_1 + 0$$

In the above models, *y* is the dependent variable, x_1 and x_2 are independent variables, *a,* and *b* are adjustable parameters, and *e* is an error term to correct for observational errors. FIT is a *fitted model* where all the parameters are fixed. M1 and M2 represent families of curves or fitted models. For example, M1 represents all the curves that could occur when values are applied to the parameters. Note that FIT is a member of the family of fitted models of M1 (and M2).[245] The dependent variable is the measurable quantity of interest, and the independent variables are the causes that influence that quantity.

Model selection is concerned with fitting models to data, a process called curve fitting. Consider a graph of collected data. From a realist perspective, it is assumed that there is a true curve that generates the data (give or take observational error). The goal of model selection is to find a model that is as close to the true curve as possible given the available data. Practicing scientists know that when the data set is small, simpler models tend to be better predictors than more complex models. In fact, it is well known that curves that perfectly go through every data point tend to be poor predictors because they overfit the data. The problem with overfitting is that it mistakes observational error for as a true cause of the target system. If the goal of scientific realism is to discover true causes, and model selection can be used to identify true causes of a target system by avoiding overfitting and increasing PA, it may be possible to use model selection criteria to overcome the problems of verisimilitude.

Predictive Accuracy and AIC

Predictive accuracy, as defined by Forster and Sober (1994), is the ability for a selected model to predict new data given existing data. In situations where there is little data available, a simple model may be more predictively accurate than a more complex one, but as more data becomes available, the choice of models may be revised because the simpler model fails to be as predictively accurate. For example, in data poor situations, a simple model like M1 may be more predictively accurate, but, as the amount of

[245] All of the fitted models of M1 are within the family of M2 where *b* equals 0.

data increases, a more complex model like M2 may be selected because of its greater ability to predict new data.

Although there are many types of model selection theories, this paper is concerned with AIC due to its relation to verisimilitude. Forster (2000) explains that an important part of AIC is that "the conclusions of AIC are . . . about its closeness to the truth" (213). If the true curve is maximally predictively accurate, and if AIC chooses the maximally predictively accurate curve given the data available, increasing PA can overcome the logical problem and AIC should overcome and epistemic problem.

The purpose of AIC is to minimize the Kullbach-Leibler distance[246] (K-L) between potential fitted curves within a family and the true curve represented by the data (Forster 2000). K-L distance, as defined by Burnham and Anderson (2002), indicates the distance between a candidate model and the true curve. However, since K-L distance cannot be computed without a prior knowledge of the true curve, a selection criterion like AIC must be used (Burnam and Anderson 2002). AIC, then, is supposed to provide an estimation of the closeness to the truth of a model, *M*. Sober (2008) provides the following formulation of AIC:

$$AIC(M) =_{df} Log\{\mathrm{Pr}[Data|L(M)]\} - k$$

In this formulation, *L(M)* represents the likeliest fitted model of *M* given the data available. *AIC(M)* is found by taking the log likelihood of *L(M)* and subtracting a penalty for complexity, *k*. The term *k* represents the number of parameters in the model and is used to prevent AIC from overfitting a model given the data when models are being compared. Complex models always fit the data better than simpler models, but as noted earlier, complex models are not always better predictors due to problems of overfitting. By having the correction for complexity, AIC is able to provide a reliable estimate of the model's PA. Thus, AIC only selects a model with a greater number of parameters when the log likelihood overcomes the *k* penalty.

Because AIC scores are dependent on the size of the data set, as the amount of data increases, AIC could select more complex models. For example, assume that there are three candidate models:

$$(M1) \quad y = ax_1 + e$$

$$(M2) \quad y = ax_1 + bx_2 + e$$

$$(M3) \quad y = ax_1 + bx_2 + cx_3 + e$$

In a data poor situation, AIC might favor the simpler model such that the following inequality holds: *AIC(M1)>AIC(M2)>AIC(M3)*. As we gather more evidence and the size of the data set increases, the AIC

[246] It is worth noting that the K-L distance is not a true distance because it does not satisfy the triangle inequality. However, for the purposes of this paper the term "distance" works to clearly relate the concept of closeness or proximity between curves.

might recommend *M2* over *M1* if the AIC score of *M2* is greater than *M1*. If it is true the x_2 is a new cause affecting the system, then it may seem that increasing PA will likewise increase closeness to the truth. In this way, the use of PA and AIC makes great progress at dealing with both the logical and epistemic problems. Forster and Sober (1994) indicate that minimizing K-L distance to the true curve is the same as maximizing predictive accuracy. When selecting a model with the best AIC score, the model being selected is the closest model to the true curve given the available data.

The contrastive nature of PA and AIC also seem to overcome the epistemic problem that PCV failed to do. As new data is gathered, AIC may select a different family of curves with greater predictive accuracy than the current model. Because there is an existing metric of truth with the AIC score, obtaining a better score and increasing PA provides a contrastive view of progress similar to what Popper had attempted to do with PCV. In the examples of *M1, M2,* and *M3* above, when AIC selects *M2* over *M1,* an increase in closeness to the truth is being made along with an increase in predictive accuracy. That is, the new model is capturing more true causes of the target system while increasing the ability to accurately predict new data.

When AIC Fails

However, the ability for PA and AIC to overcome the logical and epistemic problems is based on ideal data situation. In data poor or data rich situations, there are complications that arise and create interesting dilemmas. Assume, for example, there is a target system that has three causes previously identified; however, the size of the data set is small. Even though we may know there are three causes of the target system, AIC may select a simpler model with only one cause because it will have greater predictive accuracy instead of a model that includes all three causes and is closer to the truth. This wrinkle may seem minor, but it shows that AIC may be tracking our ability to predict new data rather than tracking a theory's closeness to the truth in such a way that, while it can overcome the logical problem, it only does so in ideal data situations. However, the epistemic problem is still answered since, as data increases, AIC select models that do identify more true causes of the target system as the predictive accuracy increases for those models.

Before turning to the next dilemma, the *e* term, observational error, must be discussed. All of our scientific inquiry is subject to observational error or noise that is included in a data set. AIC assumes that observational error is present and accounts for it, but the very presence of observational error is what leads to the greater problem behind AIC. As such, there is a possibility that AIC will fail in data rich situations by selecting models that are further from the truth. While the error term included in models is supposed to deal with observational errors, as data sets get larger, there is a chance that AIC will recommend an additional parameter that is not a cause of the system being investigated. In other words, our model selection framework might be tracking the cause of observational error and mistakenly attributing it as a cause of the system under investigation. Forster and Sober (1994) explain that AIC was designed to estimate the size of the overfitting factor, but they also mention that the process is fallible. Given the possibility for AIC to recommend an error term as a new cause, we are now left with an interesting dilemma wherein either the logical problem or the epistemic problem will reassert itself. I will consider each horn of the dilemma separately.

I will begin by addressing the first horn. If our goal is to discover all the true causes affecting the target system, then in data rich situations we cannot be sure that a newly discovered variable is representing a cause of the target system or a cause of our observational error. If AIC is identifying causes of something outside of the target system, then there are some cases where we cannot tell whether progress is being made even if we are increasing predictive accuracy.

To illustrate the second horn of the dilemma, we can consider how a defender of the model selection framework might reply to the first horn. One might maintain that increasing PA always gets us closer to some truth. However, the truth being identified by increasing PA ceases to be about the target system, but begins to track the truth about the system that generates the data. This new system would take account of both the target system *and* the causes of our observational error. In such a situation, we give up the noumena in favor of the phenomena – we exchange our realist notion of the truth of a target system for the appearance created by the data. It is hard to see how such a solution would be palatable to scientific realists. Since the logical problem was supposed to allow for scientific realism, it seems that such a step gives up on the logical problem altogether.

These two horns of AIC create a trade off when dealing with the logical and epistemic problems. Either we accept that our choice in models can select better theories but we cannot always tell if we are getting closer to the truth, or we give up on scientific realism in favor of the notion that models with greater PA are closer to the truth about the system that gives rise to the data but not the true target of our inquiry.

Conclusion

PA and AIC seem to be heading in the right direction in understanding progress. However, if providing answers to Popper's logical and epistemic questions are the criteria by which a true sense of progress can be determined, PA and AIC seem to fall short of the mark if we want to maintain a realist approach to progress in all cases. The problem of data poor situations can be overcome by increasing the size of the data pool, and progress towards the truth can still be made. However, in data rich situations that may not be the case. Although AIC runs into this problem at the extreme limit, and it's likely that our extant theories have yet to run into it, there is a possibility that AIC will stop modeling the true causes of the target system at some point, and increasing PA will no longer be progress towards the truth of the target system. Of course, increasing PA and selecting a model with the best AIC, in ideal data situations, does seem to satisfy both the logical and epistemic problem, so it may give progress hope. In terms of theories that can capture closeness to the truth and the movement of progress, PA and AIC seem to come closer than Popper's first attempt. Reminiscent of Popper's hypothetico-deductive method, PA and AIC seem to hold up to more severe tests than Popper's theory of verisimilitude did, and, in some ways, that seems like it should be progress in itself.

Bibliography

Chakravartty, Anjan. "Scientific Realism." *The Stanford Encyclopedia of Philosophy. Last modified 2015.* http://plato.stanford.edu/archives/fall2015/entries/scientific-realism.

Forster, Malcolm. “Key Concepts in Model Selection: Performance and Generalizability.” *Journal of Mathmatical Psychology* 44 (2000): 205-231, http://philosophy.wisc.edu/forster/papers/JMP2000.pdf.

Forster, Malcolm. “Verisimilitude and Likelihood.” Manuscript, University of Wisconsin: Madison, 2004.

Forster, Malcolm Elliot Sober. “How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions.” *The British Journal for the Philosophy of Science* 45, No. 1 (1994): 1- 35, http://www.jstor.org/stable/687960.

Goldsby, Michael. “The “Structure” of the “Strategy”: Looking at the Matthewson-Weisberg Trade-off and Its Justificatory Role for the Multiple-Models Approach.” *Philosophy of Science* 89, no. 5 (2013): 862-873, http://www.jstor.org/stable/10.1086/673728. .

Miller, David. “Popper’s Qualitative Theory of Verisimilitude.” *The British Journal for the Philosophy of Science* 25, No. 2 (1974): 166-177, http://www.jstor.org/stable/686821.

Oddie, Graham. "Truthlikeness." *The Stanford Encyclopedia of Philosophy. Last modified 2014. http://plato.stanford.edu/entries/truthlikeness.*

Popper, Karl. *Conjectures and Refutations*. London: Routledge, 1963.

Popper, Karl. *The Logic of Scientific Discovery.* New York: Routledge, 1959.

Sober, Elliot. *Evidence and Evolution: The Logic Behind the Science.* Cambridge: Cambridge University Press, 2008.

Tichý, Pavel. “On Popper’s Definition of Verisimilitude.” *The British Journal for the Philosophy of Science* 25, No. 2 (1974): 155-160, http://www.jstor.org/stable/686819.

Van Fraassen, Bas. *The Scientific Image.* Oxford: Oxford University Press, 1980.