A Gruesome Problem for the Curve-Fitting Solution

Scott DeVito
DISCUSSION

A Gruesome Problem for the Curve-Fitting Solution
Scott DeVito

ABSTRACT

This paper is a response to Forster and Sober's [1994] solution to the curve-fitting problem. If their solution is correct, it will provide us with a solution to the New Riddle of Induction as well as provide a basis for choosing realism over conventionalism. Examining this solution is also important as Forster and Sober incorporate it in much of their other philosophical work (see Forster [1995a, b, 1994] and Sober [1996, 1995, 1993]). I argue that Forster and Sober's solution is subject precisely to the problem they seek to solve. They provide a method of choosing among hypotheses but only at the cost of requiring that we have a method of choosing between different ways of conceptualizing the world. Thus the solution raises a new problem—the world-fitting problem.

1 Hypothesis choice and Akaike's theorem
2 A 'Gruesome' problem
3 An everyday problem
4 Conclusion

1 Hypothesis choice and Akaike's theorem

The curve-fitting problem arises when we have a set of data that can be modelled equally well by more than one curve. When such a situation arises, the data, by themselves, cannot be used to choose among these curves. In addition, each curve represents a different hypothesis about the world and the curves are mutually exclusive. The curve-fitting problem is the problem of devising a way of choosing a curve, from among these mutually exclusive and equally well-fitting curves, that leads to our choosing the curve that is closest to the true curve.

An important measure of a curve's closeness to the truth is the curve's predictive accuracy. One might think that a solution to the curve-fitting problem would be one in which we choose the particular curve that has the highest predictive accuracy. But such an approach with its focus on particular curves is open to considerable problems. Forster and Sober attack the curve-fitting problem somewhat differently. They use Akaike's theorem and its

© Oxford University Press 1997
notion of a family of curves' predictive accuracy. Forster and Sober's solution to the curve-fitting problem provides us with a way to choose one curve from among a set of curves where each curve is from a different family of curves ([1994], pp. 4 and 11).

Akaike's theorem can be stated as follows:

\[
\text{Estimated}[A(\text{family } F)] = \frac{1}{n} \log(\text{likelihood}(L(F)) - k)
\]

where \(A(\text{family } F)\) is the measure of the accuracy of the family \(F\), \(k\) is the number of adjustable parameters for the family \(F\), \(L(F)\) is the curve in \(F\) that best fits the data, and \(n\) is the number of data points.

Simplicity is defined relative to the number of adjustable parameters of a family of curves. The family of curves \(F_1\) is simpler than the family of curves \(F_2\) just in case the number of adjustable parameters in \(F_1\) is less than the number of adjustable parameters in \(F_2\). The standard examples are of the family of straight lines, LIN (which contains curves of the form \(y = a + bx\)), and the family of parabolas, PAR (of the form \(y = a + bx + cx^2\)). Since LIN has two adjustable parameters (\(a\) and \(b\)) it is simpler than PAR, which has three adjustable parameters (\(a\), \(b\), and \(c\)).

2 A 'Gruesome' problem

Akaike's theorem does not provide a worry-free method for choosing between hypotheses. The problem with using Akaike's theorem for hypothesis choice is that the number of parameters associated with a given hypothesis is a matter of convention. In addition, for any hypothesis there is no a priori way to generate the right family of curves to which the hypothesis belongs.

This problem is well established in the literature on the new riddle of induction. We can see how the problem plays out by attempting to apply Akaike's theorem to choose between two hypotheses: the Grue hypothesis and the Green hypothesis.

**Grue hypothesis:** If \(E\) is an emerald and it is before the year 2100 \((t < 2100)\), then \(E\) is green; otherwise, if \(E\) is an emerald and \(t \geq 2100\), then \(E\) is blue.

**Green hypothesis:** If \(E\) is an emerald then it is green.

The question we are concerned with is, given that these two hypotheses fit the data equally well and that they are mutually exclusive, how do we choose between them?

According to Forster and Sober, Akaike's theorem provides a solution to this problem. Since both hypotheses fit the data equally well, Akaike's theorem

---

1 A family of curves is a set of curves whose equations can be obtained from a given equation by varying \(n\) essential constants (adjustable parameters) which occur in the given equation.
tells us that we should choose the hypothesis that comes from the simpler family. The Grue hypothesis has one adjustable parameter, the time $t$. The Green hypothesis has no adjustable parameters. So, the Green hypothesis is simpler than the Grue hypothesis and should be chosen over the Grue hypothesis.

But as is well known, it is possible to restate both the Grue hypothesis and the Green hypothesis so that their respective number of adjustable parameters is switched—the Grue hypothesis would have no adjustable parameters while the Green hypothesis would have one adjustable parameter. (And both modified hypotheses would still fit the data equally well.)

To do so we merely have to take grue and bleen to be our basic colour terms. (In the description above green and blue were taken to be our basic colour terms.) In essence, we change the way we think of or conceptualize the world. Things in the world are no longer taken to be green or blue; they are thought to be grue or bleen.

In a world where we replace green and blue with grue and bleen we would state the Grue and Green hypotheses in a different manner.

**Grue hypothesis:** If $E$ is an emerald, then $E$ is grue.

**Green hypothesis:** If $E$ is an emerald and $t < 2100$, then $E$ is grue; otherwise, if $E$ is an emerald and $t \geq 2100$, then $E$ is bleen.

If we conceptualize the world in this manner, then the Grue hypothesis has no adjustable parameters and the Green hypothesis has one, the time $t$.

Since both hypotheses fit the data equally well, are mutually exclusive, and are from different families of curves, Akaike’s theorem tells us to choose the hypothesis from the simpler family of curves. Since the Grue hypothesis, in this situation, comes from the simpler family of curves we should choose it over the Green hypothesis.

This result raises a problem for using Akaike’s theorem as a way to choose between competing hypotheses. The results Akaike’s theorem gives are relative to the way that we conceptualize the world. If we adopt green and blue as our basic colour terms, then Akaike’s theorem tells us to choose the Green hypothesis over the Grue hypothesis. But if we adopt grue and bleen as our basic colour terms, then Akaike’s theorem tells us to choose the Grue hypothesis over the Green hypothesis. As Sober (1994) has previously argued, ‘no model, no inference’.

---

2 This example/idea is similar to an example of Gilbert Harman ([1994], p. 159).

3 Forster (in private correspondence) is concerned that I have duplicated the work of Priest [1976] and that Priest’s work does not provide an objection to the use of Akaike’s theorem to solve the curve-fitting problem. Forster and Sober take Priest to be defining simplicity as relative to a particular curve. Since Forster and Sober define simplicity relative to a family of curves, they believe that they avoid the objections raised by Priest to using simplicity as part of the solution to the curve-fitting problem (Forster and Sober [1994], p. 11). Since I focus exclusively on the simplicity of families of curves, my objection cannot be rejected by Forster and Sober in this manner.
3 An everyday problem

This problem does not only arise when we have ‘cooked up’ hypotheses like Grue and Bleen; it arises for any standard curve. Take the data set D, where the curve in family of curves LIN (lines) that best fits the data is $H_1 = a_0 + a_1 x$ and the curve in family of curves PAR (parabolas) that best fits the data is $H_2 = b_0 + b_1 x + b_2 x^2$. Let us assume that the log-likelihood of $H_1$ equals the log-likelihood of $H_2$. Then, Akaike’s theorem tells us to choose $H_1$ over $H_2$ because LIN is simpler than PAR.

As we noted above, the family of curves to which a curve belongs (and its associated simplicity measure) can change if we alter our background assumptions about the world—if we change the way that we conceptualize the world.

$H_1$ and $H_2$ were originally compared in the standard Cartesian coordinate system $[X,Y]$. If we change the coordinate system to $[X', Y]$, where $X' = x' + b_2 x^2$, we find that $H_2$ is preferred over $H_1$. In $[X', Y]$, there is a family of curves PAR', to which $H_1'$ ($H_1$) belongs, and a family of curves LIN', to which $H_2'$ ($H_2$) belongs. $H_1'$ is the member of PAR' that best fits the data and $H_2'$ is the member of LIN' that best fits the data. The form of the members of PAR' is $\gamma_0 + \gamma_1 x'$ and the form of the members of LIN' is $\gamma_0 + \gamma_1 x'$. Since both $H_1'$ and $H_2'$ fit the data equally well and $H_1'$ is a member of a more complex family of curves than $H_2'$, Akaike’s theorem tells us to choose $H_2'$ over $H_1'$. But, this result is inconsistent with the result achieved in the $[X,Y]$ coordinate system. Once again Akaike’s theorem leads to different results when we compare hypotheses using different conceptualizations of the world.

4 Conclusion

Forster and Sober have argued that they have a solution to the curve-fitting problem. As we have seen above, Forster and Sober’s solution is subject to a serious problem—it cannot provide us with a method of choosing between hypotheses except in the context of a particular conceptualization of the world. If we adopt a particular conceptualization of the world, then Akaike’s theorem gives us a way to choose between hypotheses under that conceptualization of the world. If we conceptualize the world differently, then Akaike’s theorem might tell us to choose a different curve.

4 For ease of reading, and simplicity of argumentation, I assume that the log-likelihood of $H_1$ equals the log-likelihood of $H_2$. This assumption is not necessary. The arguments presented in this paper work just as well if we assume that the difference in the log-likelihood of $H_1$ and the log-likelihood of $H_2$ is small.

5 For $H_1'$ and $H_2'$ to fit the data equally well, we have to assume that the compression of the data points along the X' axis (as compared to the spread of data points along the X-axis) does not make $H_1'$ closer to the data than $H_2'$ or make $H_1'$ closer to the data than $H_1'$. Since this is not logically impossible, it takes the assumption to be plausible.
A Gruesome Problem for the Curve-Fitting Solution

Forster and Sober are realists. As such, they want a realist solution to the curve-fitting problem (see e.g. Forster and Sober [1994], p. 28). If Akaike’s theorem tells us that we should choose one curve over another, then the chosen curve should be closer to the truth than the rejected curve. Unfortunately, the results Akaike’s theorem provide are relative to the conceptualization of the world within which we compare the hypotheses. Akaike’s theorem does not provide a realist solution to the curve-fitting problem.

But all is not lost. We can still adopt Akaike’s theorem (or something like it) as a method for choosing between hypotheses. To do so we must accept some form of conventionalism (or perhaps some type of Bayesianism). To apply Akaike’s theorem we need to specify the conceptualization of the world that we have adopted (or the background assumptions we have about the world). Once we have done so, we can use Akaike’s theorem to choose between hypotheses. At the same time, we must recognize that the choice is conceptualization-specific and does not guarantee that we have chosen the curve closest to the truth.

Acknowledgements

I am greatly indebted to Prasanta Bandyopadhyay, John G. Bennett, Malcolm Forster, Henry E. Kyburg, Jr, and Andrew Wayne for their insightful discussion and comments on earlier drafts of this paper.

Department of History and Philosophy of Science
1017 Cathedral of Learning
University of Pittsburgh
Pittsburgh, PA 15260
USA

References


