Studies in History and Philosophy of Biological and Biomedical Sciences 60 (2016) 35–43

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

AIC and the challenge of complexity: A case study from ecology

Remington J. Moll [a,*], Daniel Steel [b], Robert A. Montgomery [a,c]

[a] Department of Fisheries and Wildlife, Michigan State University, 480 Wilson Road, 13 Natural Resources Building, East Lansing, MI 48824, USA
[b] The Maurice Young Center for Applied Ethics, School of Population and Public Health, University of British Columbia, 227 – 6356 Agricultural Road, Vancouver, BC V6T 1Z2, Canada
[c] Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, Recanati-Kaplan Centre, Tubney House, Abingdon Road, Tubney, Oxfordshire OX13 5QL, UK

ARTICLE INFO

Article history:
Received 10 March 2016
Received in revised form
9 September 2016
Available online 30 September 2016

Keywords:
Akaike's Information Criterion
Complexity
Ecology
Modeling
Model selection
Simplicity

ABSTRACT

Philosophers and scientists alike have suggested Akaike's Information Criterion (AIC), and other similar model selection methods, show predictive accuracy justifies a preference for simplicity in model selection. This epistemic justification of simplicity is limited by an assumption of AIC which requires that the same probability distribution must generate the data used to fit the model and the data about which predictions are made. This limitation has been previously noted but appears to often go unnoticed by philosophers and scientists and has not been analyzed in relation to complexity. If predictions are about future observations, we argue that this assumption is unlikely to hold for models of complex phenomena. That in turn creates a practical limitation for simplicity's AIC-based justification because scientists modeling such phenomena are often interested in predicting the future. We support our argument with an ecological case study concerning the reintroduction of wolves into Yellowstone National Park, U.S.A. We suggest that AIC might still lend epistemic support for simplicity by leading to better explanations of complex phenomena.

1. The epistemic value of simplicity in modeling

Philosophers and scientists have long valued simplicity, but the epistemic justification of this value has been intensely debated. Over the past two decades, one facet of the debate has focused on statistical model selection. In a 1994 article, Forster & Sober proposed that Akaike's Information Criterion (AIC) provided such a justification by linking simpler models to increased predictive accuracy. Forster & Sober's proposal has been influential in the philosophical literature,[1] and their analysis has been extended to topics such as predictivism (Dowe, Gardner, & Oppy, 2007; Forster, 2002; Hitchcock & Sober, 2004; Lee, 2013; Sober, 2008). The article has also enjoyed uptake in the scientific literature (e.g., Ginzburg & Jensen, 2004) and influential statistical texts (e.g., Burnham & Anderson, 2002). The proposal has also been criticized on the grounds that AIC is not invariant under changes in how models are described or grouped into families (DeVito, 1997; Forster, 1995, 1999b; Kukla, 1995). We raise a distinct and largely undiscussed concern that Forster and Sober's proposal is limited when modeling complex phenomena because complexity challenges an assumption required by AIC. Our argument has important implications because AIC is routinely used by modelers studying complex phenomena.

Forster & Sober note that although models can be formed and parameters estimated such that a model passes exactly through every data point, it is common practice to avoid this technique. This is because "scientists seem to be willing to sacrifice goodness-of-fit if there is a compensating gain in simplicity" (Forster & Sober, 1994, p. 5). Techniques that maximize goodness-of-fit without considering simplicity (e.g., ordinary least squares regression) are sometimes insufficient, as we shall see in Section 2. Moreover, a vague or subjective notion of simplicity will undermine its epistemic justification. What is required is a criterion that trades-off fit with simplicity in an empirically justifiable fashion. Forster and Sober (1994) argue AIC does exactly this because it improves predictive accuracy by balancing models' simplicity and fit to the data. Drawing a link between AIC and predictive accuracy requires the assumption that the data used to fit the model and the data for which predictions are made are generated by the same probability distribution (Forster, 2000, p. 216). Following Forster & Sober, we

* Corresponding author.
E-mail address: rjmoll@msu.edu (R.J. Moll).

[1] As of August 16, 2016, Google Scholar lists 455 citations of the article.

http://dx.doi.org/10.1016/j.shpsc.2016.09.007

call this the "uniformity of nature" assumption (Forster & Sober, 1994, p. 29; Sober, 2002, p. S116).[2]

The uniformity of nature assumption means that Forster & Sober's philosophical justification of simplicity is limited to what Forster (2002) calls interpolative predictive accuracy, that is, predictive accuracy for the data to which a model was fit; it does not extend to extrapolative predictive accuracy. Although mentioned by Forster and Sober (1994) and discussed by Forster (2000, 2002), this limitation appears to often go unnoticed by both philosophers and scientists. Clarifying AIC's relationship to extrapolative predictions is crucial because scientists typically understand predictions to be about the future and predictive accuracy in this sense is often cited as a goal of AIC (Aho, Derryberry, & Peterson, 2014; Shmueli, 2010). In addition, neither AIC nor Forster & Sober's justification of simplicity have been philosophically analyzed in relation to modeling complex phenomena, which is surprising given that AIC is routinely used by modelers working in complex disciplines such as ecology. Here we use an ecological case study to show how complexity challenges the uniformity of nature assumption when model predictions are extrapolated. This case clarifies how AIC tends not to increase predictive accuracy as many scientists understand it. We offer the alternative idea that when AIC is applied to models of complex phenomena, it serves to improve explanations of past events rather than improve predictions of future ones. Our argument is relevant not only to ecology, where the use of AIC is pervasive (Burnham & Anderson, 2002),[3] but also to similar disciplines that apply AIC such as sociology (Burnham & Anderson, 2004) and psychology (Wagenmakers & Farrell, 2004).

The organization of this article is as follows. In Section 2, we describe AIC and Forster & Sober's philosophical justification of simplicity. In Section 3, we use an ecological case study to demonstrate how complexity threatens the uniformity of nature assumption. In Section 4, we respond to potential objections to our argument. In Section 5, we explore the idea that AIC can lead to better explanations. We conclude our argument in Section 6.

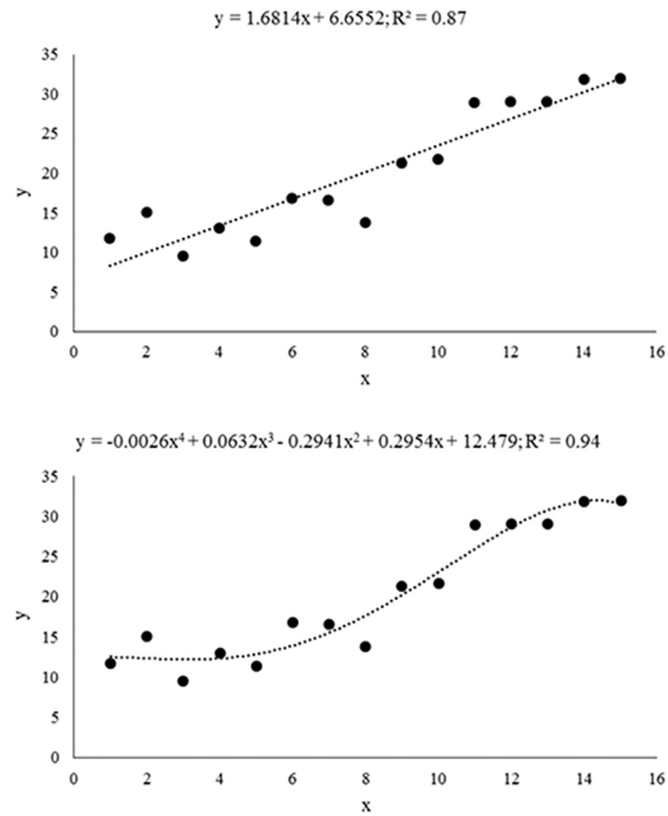

Fig. 1. A linear process generates the data (the true model is $y = 1.6x + 6$), but observed data include error (normally distributed with a mean of 0 and a standard deviation of 3). Two models are fit using ordinary least squares: top, a linear function, and bottom, a fourth-order polynomial function. The polynomial fits the data better (as indicated by the R^2 statistic), but will be a poorer predictor of future data from the (true) linear process.

2. AIC and the uniformity of nature assumption

Our discussion turns on the question: how should models be fit to data? Model fitting has two components. First, a model form must be chosen. The linear model represents the 'base' or simplest model form[4]:

$$y = ax + b + \varepsilon, \tag{1}$$

where y is some response variable, x is some predictor variable, a is the slope, b is the y-intercept, and ε is an error term. Models with higher-level polynomials (e.g., add x^2 to Eq. (1)) or additional predictors (x_1, x_2, etc.) are considered more complex than the linear model. The choice among competing models is termed *model selection*. Second, values of a given model's parameters must be estimated. For example, a and b in Eq. (1) must be somehow optimized because there are infinitely many straight lines that could be fit to the data. The process of *parameter estimation* seeks those values that make a given model's predictions most closely match the data. A simple approach is ordinary least squares, which minimizes the squared difference between model predictions and the data. The maximum likelihood method of parameter estimation is a generalization of this approach (Burnham & Anderson, 2002) and differs in that, instead of squared differences, a likelihood function is used to relate the model to data.[5]

Forster and Sober (1994) note that the naïve empiricist should unequivocally select a more complex model because it will always minimize the error between the model and the data. To choose a simpler model that fits the data less well as a mere matter of taste (cf. Kuhn, 1957, p. 181) would be foolhardy from an empiricist perspective. The empiricist could prefer the simpler model only if it provided an epistemic advantage, such as increased predictive accuracy.

Forster & Sober use the concepts of signal and noise to explain how a simpler model might make better predictions. The signal is the true process generating the data and the noise is the observational error in the data collection process. If we perfectly fit a complex model to the data, we conflate the signal with the noise and thereby decrease predictive accuracy due to Type I errors (i.e., finding spurious relationships among variables; Freedman, 1983). Forster & Sober provide an example to support their claim. Suppose

[2] The label "uniformity of nature" is potentially misleading. In the present context, it only refers to the existence of a single probability density function that generates data throughout the time period of concern. It is not a claim about the existence of stable, universal physical laws.

[3] This authoritative reference on AIC has 33,413 citations on Google Scholar as of August 16, 2016.

[4] In applied sciences, the term 'linear model' can refer to models for which the response variable is predicted as a function of predictor variables with additive effects. In Eq. (1), a 'linear model' means a straight line.

[5] Bolker (2008) describes this as finding the parameter values that, given the model, make the data the "most likely to have happened" (p. 170), but it's important to emphasize that a likelihood function considers the data fixed and the parameters variable, not the other way around.