Ockham's Razor

Seminar Wissenschaftliche Methoden, 24.06.11
Laurin Ostermann

1 General Overview

The role of Ockham's Razer in the philosophy of science is to give an idea or guideline to which theory to root for, amongst a pool of several, if they all possess the same congruence with gathered data: "From theories fitting the data equally well, scientists should choose the simplest one".

Yet, most accurate fit to given data is not the only criterion constituting a good theory choice. Parsimony, predictive accuracy, *explanatory power*, fruitfulness in explaining new insights, *testability*, *repeatability*, consistence with other scientific and philosophical beliefs, ... should also be considered.

Various synonyms basically refer to the mentioned concept: parsimony, principle of simplicity, principle of economy and Ockham's Razor.

Parsimony as a ground-rule is important in every scientific field, strengthening the foundation of any theory. It is cross-disciplinary.

Parsimony is important because science cannot and will not produce any single conclusion without invoking parsimony. Every logical inference will at some point disregard some (minor) details in order to come up with a more general statement, reducing a complex matter down to a few simple principles. Furthermore, parsimonious models can be much more effective in terms of data collection or computational time, which is almost always directly or indirectly connected to a monetary advantage: "Scientists want to find the truth, but don't want to spend more time or money than necessary."

2 Historic Perception

There are two different, but related meanings. On the one hand, parsimony is an *ontological* principle, claiming that nature always chooses the simplest course. On the other hand, it is also an *epistemological* principle, suggesting that scientists should choose the simplest theory that fits the data.

From the ancient Greek philosophers to modern-day statisticians, parsimony has been regarded as a significant property of scientific theories. In the following a few historical instances in various schools of thought throughout the centuries shall be examined.

- **Aristotle** "We may assume the superiority ceteris paribus [over other equal things] of the demonstration which derives from fewer postulates or hypotheses.", and "The principles should, in fact, be as few as possible, consistently with proving what has to be proved."
- **Claudius Ptolemey** in Almagest (gold-standard astronomy book throughout the middle age) invoked parsimony to help decide between theories of planetary motion.
- **Thomas Aquinas** "If a thing can be adequately done by means of one, it is superfluous to do it by means of several; for we observe that nature does not employ two instruments where one suffices."
- **Robert Grosseteste** also emphasised parsimony, held it as a real objective principle of nature more than as a criterion of good explanations.

An important role in this historic discussion must be attributed to **William of Ockham**, a medieval scholar, probably best known for his principle of parsimony, *Ockham's Razor*: "Plurality is not to be posed without necessity." (*"pluralitas non est ponanda sine necessitate."*) and further, "What can be explained by the assumption of fewer things is vainly explained by the assumption of more things.". The common form of this principle, "Entities must not be multiplied without necessity." (*"entita non sunt multiplicanda sine necessitate"*), does not seem to be his actual phrasing.

Ockham demands that "everyone who makes a statement must have a significant reason for its truth", where a significant reason can be an *observation of a fact* , an *immediate logical insight* or a *divine relation* or a deduction from there. Though experience seems to justify plurality, "one should not complicate explanations where simple ones will suffice" or, to be more dogmatic, "prefer the simplest model that fits the data accurately".

Ockham insisted on parsimony to be an epistemological principle,, in contrast to his predecessor R. Grosseteste or his teacher John Duns Scotus. "To insist that nature always takes the simplest path is to limit God's power". Therefore, he shifted simplicity from the coursed of nature to the theories which are formulated about it.

A striking example is Ockham's rejection of the impetus theory of motion (Jean Buridan, based on Aristotle) [Impetus: a force a mover embeds into an object when setting it in motion]: "Motion is a concept, having no reality apart from moving bodies, to describe the fact that from instant to instant a moving body changes its spacial relation to some other body without intermediate rest. There is no need to postulate any external or internal efficient cause to explain such a sequence of events. “Motion is neither a separate thing, nor a property of a thing, but rather a modification of existing things, namely, a change of location over time.”. This idea lead to the development of the impulse concept in classical mechanics in the 17th century.

Having devoted a closer treatment to William of Ockham, let us return to more instances of parsimony throughout the history of science.

- **Nicolaus Copernicus** brought forward the heliocentric cosmology as opposed to the - at that time - established geocentric one, mainly based on arguments in the realm of parsimony. Putting the sun in the center with planets, including the earth, orbiting around it, allowed for a description relying on less cycles and epicycles. With this he could also unify parameters, bringing together the characteristic times for the sun, mercury and venus. The experimental accuracy did not suffice to justify any of these claims. Copernicus based this concept purely on the grounds of simplicity. Bessel, through parallax measurements, finally confirmed the hypothesis - over a century later.

For balance, a quite prominent example of parsimony gone wrong shall also be mentioned.

- **Galileo Galilei** tried to unify the Aristotelian dynamics, which claim that there are two types of motion: outside the sphere of the moon only circular motion is possible and below the moon sphere there is only rectilinear motion with heavy bodies moving downwards to the center of the earth and lighter -than-air bodies moving straight up. Galileo claimed tat the latter is an illusion and, in fact, only circular motion exists: a ball falling from a tower follows a curved trajectory when observed from off the earth. [Tough this is certainly wrong, the underlying thinking of a unified theory of motion is very comprehensible and realised nowadays in the gRT.]

Moving forward,

- **Isaac Newton** formulated parsimony as an ontological and epistemological concept as the first two rules in his *philosophiae naturalis principia mathematica*, his magnum opus.

- **Gottfried W. Leibniz** proved that the path difficulty of a light ray (geometric path times resistance of the medium) is a minimum, using his differential calculus.

- **Albert Einstein** after having found the Einstein equations in the gRT concluded: “God would not have passed up the opportunity to make nature that simple.". Also, "Everything should be made as simple as possible, but not simpler.".

More recently, statisticians also observed two things: First, simple theories tend to make more reliable predictions (Bayesian statistics gives simple theories the highest prior probabilities; resulting posterior probabilities will also be higher for simpler theories). And secondly, there is also a considerable gain in accuracy and efficiency.

In addition, nowadays with sensible computing power at hand, parsimonious models are the easiest to test, to implement, etc. and also use the least calculation time.

Parsimony in general can be understood as an instance of a larger entity: beauty (compact formulas, clear pictures, comprehendible explanations, etc.)

In conclusion, history has shown that many paradigm shifts were brought about on the basis of parsimony, much rather than inspired by new measurements. Also, false theories would rather conflict with parsimony, than with data (e.g. Copernicus/Bessel). The lesson that can be gained from this is, that employing parsimony alongside theories and data, most likely puts a scientist to the cutting edge of his field.

3 Principles & Example

Let us now review a few basic principles and thereby explain the lines in fig. 1.

1. **Signal and Noise**
   At first we have to address the issue why the two curves corresponding to signal and noise look differently: signal typically has a few simple causes and can be captured by a relatively parsimonious model, while noise has manny a different origins and is usually, of corse, not reproducible. So, therefore the initial focus on signal suppresses the noise, yet, after most of the signal has been captured, noise starts to be picked up as well.

2. **Population and Sample**
   A crucial distinction has to be made between population and sample. A sample is a sub-group of a population which is subject to some sort of experiment, where the outcome ought to be generalised and is desirably applicable to the entire population.

3. **Prediction and postdiction**
   Prediction, "population-diction", is concerned with model-building

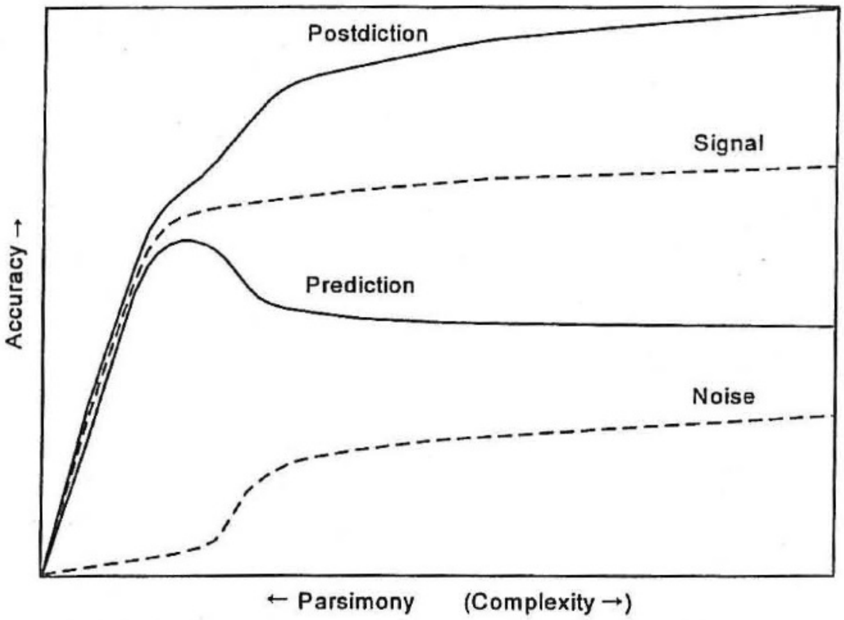

Figure 1: Accuracy vs. Parsimony (from ref. [1])

and the goal is to fit the population (future measurements, etc.), while postdiction, "sample-diction", tries to explain the data at hand. Noise has no predictive value, hence predictive accuracy is helped by capturing signal and hindered by capturing noise, therefore the prediction-line is (signal – noise). In increasing postdictive accuracy, however, we do not care about the distinction between signal and noise, thus the postdiction curve is accounted for by (signal + noise). The really important thing to note here is that most accurate *prediction*, not postdiction, is the goal in science.

4. **The Curve-Fitting Problem**
On the left-hand side of fig. 2 we find a noiseless data set and three possible model choices: two linear correlations A and B, and one of higher order, denoted by C. Almost intuitively, we would prefer line B, mainly out of a parsimony argument, since C also fits all data points. The simpler model seems more likely to predict correct values in the future. If a model with less parameters than data points fits some data set, it has already had some success. Three main advantages of a simpler model shall be named: first, a simpler model is much more vulnerable to falsification. Secondly, it delivers a more comprehensible description of the underlying process and thirdly, in a family of simple models the decision to pick out one is usually easier (here only one line would fit all the data points, while a multitude of polynomials of higher

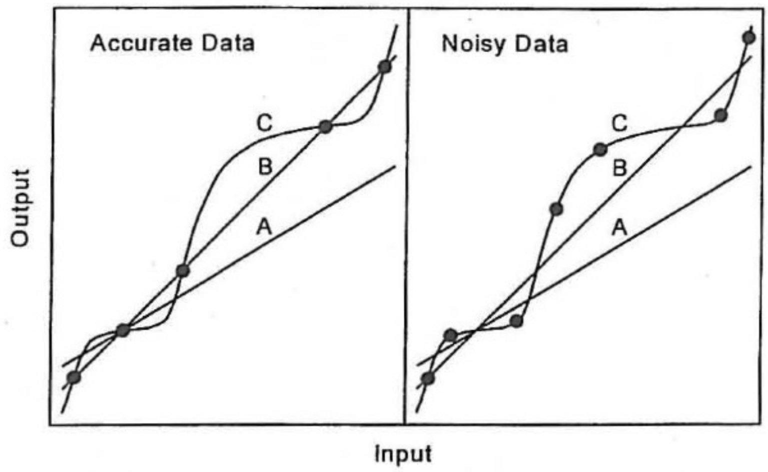

Figure 2: Curve-Fitting (from ref. [1])

order could easily fit all data points). Shifting to the picture on the right-hand side: in this case given the noisy data, one would tend to ask for more measurements. Though "experience justifies plurality", smooth and simpler behaviour have been observed much more often than abrupt changes or chaos.

5. **Related Data**
Experience shows that data concerning the same or a similar subject are usually related. This is in such a way, that given a larger data set with a few values missing, one will find his or her educated guess, inspired by the data that was there, to be not far from the actual value most of the time.

6. **Statistical Tools**
SInce a substantial mathematical treatment of statistical analysis is not what we have in mind here, we will restrict ourselves to mentioning two expressions and defining them schematically: first, the sum of squares for a data set A, where $\langle A \rangle$ denotes the grand mean of the data A, is
$$SS(A) = \sum_{a \in A} (a - \langle A \rangle)^2 \,,$$
and secondly, the statistical efficiency of a model can be determined as
$$\Gamma = \frac{SS\,(\text{data around true values})}{SS\,(\text{model around true values})}.$$
$\Gamma = 1$ means that the data is as accurate as the model, $\Gamma = n > 1$ suggests that the model is more accurate than the data and the model's

accuracy is equivalent to the accuracy one could draw from collecting n times as much data. Often also the word "signal-to-noise-ratio" shows up, even in a colloquial way. In a more rigorous statistical sense this would be $SS(\text{signal})/SS(\text{noise})$.

Let us now examine a prominent, yet easily comprehensible example of parsimony supporting scientific progress: the theory of genetics by **Georg Mendel**. He conducted experiments with peas, identifying certain properties passed on from generation to generation, where there were two different categories of properties: the ones that would immediately show up in the next generation and the ones that would be suppressed at first but show up again a few generations later down the family tree. The first ones he called *dominant* trails, while the second ones were dubbed the *recessive* ones.

Mendel treated seven different properties in his experiments which lead him to close to 1000 individual observations per trail and a ratio of dominant vs. recessive of between 2.5 and 3.2, roughly.

Invoking two parsimony arguments he was able to come up with a *theoretical* ratio. First, by implicitly postulating an underlying phenomenon, he combined the seven different trails, examining distinct properties, to one huge dominant/recessive-trial, obtaining an average ratio of 2.81 and than suggested that the ratio should be made up of small integers (simplicity and beauty), thus ending up at a result of 3:1.

With modern knowledge about DNA and the Binomial distribution this result can be confirmed, which is amazing given the fact, that the two abstractions were purely based on parsimony.

For further, more complex examples, including a precisely treatable mathematical example we refer the interested reader to ref [1].

4 Philosophic Perspective

Let us now discuss three major issues from a philosophical point of view:

1. **Parsimony vs. Accuracy Trade-Off**
The challenge here is to combine the concept of simplicity with a ceteris paribus clause about an equally good fit to the data. How do we decide the respective weights of simplicity versus fit? The discussion does not arise in dealing with postdiciton: given population data less parsimony increases the accuracy, there is no self-suggested optimum. In most of the cases, though, prediction is the goal and therefor a model at the top of Ockham's hill (fig. 1, global maximum of prediction line)

leaves us with a sensible parsimony/accuracy-trade-off, since past this maximum, neither simplicity, nor a better fit are promoted.

2. **Prediction & Truth**
   Predictive success is often taken as evidence of truth. A famous historical example of this concept is Halley's calculation of the orbit of a comet, which since then bares his name. He predicted the return of the comet correctly, relying on Newtonian mechanics and thereby accumulated acceptance for his theory of the mechanics of comets. Correct predictions are often viewed as truth, especially if a theory has proven to hold in numerous instances, so that mere luck seems very implausible. Having this association between predictive success and truth is not so much troubling for scientists, since anyway there is an advancement, but this notion is disturbing to philosophers, because frequently generic models, though not "entirely true" may suffice to generate respectable accuracy (and therefore are hard to distinguish from the true story).

3. **Parsimony & nature**
   If we incorporate parsimony in our theories (epistemological parsimony), we implicitly accept ontological parsimony, that is to say that nature is simple. "Induction, uniformity, causality, intelligibility and other scientific principles all implicate parsimony.", so "Were nature not simple, science would loose all its foundational principles at once.". And in conclusion, most importantly, "The beginning of science's simplicity is simple questions."

References

[1] H. G. Gauch, Scientific Method in Practise, Cambridge University Press, 2003

[2] E. Giannetto, The impetus theory: Between history of physics and science education, in Science & Education, Vol. 2, No. 3

[3] http://en.wikipedia.org/wiki/William_of_ockham