Open Science Framework DOI:10.31219/osf.io/***

Testable and untestable models [draft]

P.G.L. Porta Mana
Kavli Institute, Trondheim <pgl@portamana.org>

***; updated 24 May 2020

***

1 Would you like to know which one is true, eventually?

The word *model* seems to have been taking the place of *hypothesis* since around the 1960s. This replacement does not seem to be connected with the shift from frequentist to Bayesian methods.

Does the replacement of 'hypothesis' by 'model' indicate a shift in concepts in all probability theory and statistics? Or are these two words simply equivalent?

Many hodiernal authors indeed seem to use them interchangeably. An example is Kass & Raftery's famous review[1]. Initially they define a statistical model as something that represents the probability of the data according to a hypothesis (p. 773). Eventually they use the two terms interchangeably, for example saying at times 'the hypothesis H_k', at times 'the model H_k'. ***

'It's just a matter of terminology', some may say. I want to argue that it is related to something much more important: the *verifiability* of a hypothesis or model. I mean the following.

There's a closed box. You have three hypotheses: the box contains one, two, or three balls. By shaking the box a little and listening to the rattle you get the impression that there should be three balls. Maybe two. Unlikely to be just one. You may express your beliefs with a probability distribution.

But how do you *verify* how many balls there are? If possible you could open and look; or X-ray the box; or weigh it (assuming you knew the separate weights of box and balls). The electromagnetic or weigh data would tell you which hypothesis is true. You verify that there are two balls.

[1] Kass & Raftery 1995.

The verification of other hypotheses, often entertained in science, is not so straightforward. It would require an infinite amount of data or time.

Someone states that, during the whole history of a specific coin, it will come up heads 51% of the times it will be tossed, and tails 49% of the times. Another person says 50%/50%. How do you verify these hypotheses? You would need to monitor the coin, probably bequeathing this scientific task to several future generations (the oldest coin existing today is about 2 600 years old[2]), amassing a very long sequence of data. Or imagine ancient Greeks making hypotheses about the Earth's precise distance from the Sun. The data necessary to verify their hypotheses could not be gathered then, owing to lack of technology. But they were some centuries later[3]. In cases like these we say that the hypotheses are verifiable *in principle*. Scientific hypotheses are often of this kind.

There are also hypotheses that cannot be verified even with infinite data or futuristic technology. Their verification hinges on dubious notions, such as multiple copies of this universe with slightly different initial conditions. It is debatable whether we can speak of 'hypotheses' and 'verification' in this case.

The three cases above are not sharply distinct. There is an increasing difficulty with the kind or quantity of data necessary to verify a set of hypotheses.

'the null must be nested within the alternative' (end of p. 776)

I have never seen a paper in which a probability model is *refuted*. Sure, there was 'hypothesis rejection at some significance level'

***comparing two frequency-priors using exch. data is like using one data point only.

***[4] Raftery: non-testable models ***[5] is the 'hot hand' example really testable?

2 Discussion

The shift from 'hypothesis' to 'model' seems to reflect the gradual abandonment of trying to understanding a phenomenon to simply trying to fit to it some equation pulled out of a hat. This is also reflected in the defacing of the verb *explain*: many authors say 'this distribution

[2] `https://rg.ancients.info/lion/article.html`
[3] Goldstein 1985.
[4] Raftery et al. 1989.
[5] Kass & Raftery 1995.

explains the data' (or the variance of the data, or similar) when in reality it just *fits* the data – it does not explain them.

Bibliography

(‘de *X*’ is listed under D, ‘van *X*’ under V, and so on, regardless of national conventions.)

Goldstein Jr., S. J. (1985): *Christiaan Huygens' measurement of the distance to the Sun*. Observatory **105**, 32–33.

Kass, R. E., Raftery, A. E. (1995): *Bayes factors*. J. Am. Stat. Assoc. **90**^{430}, 773–795. https://www.stat.washington.edu/raftery/Research/PDF/kass1995.pdf; https://www.andrew.cmu.edu/user/kk3n/simplicity/KassRaftery1995.pdf.

Raftery, A. E., Fairley, D., Joe, H., Weissman, I., Singpurwalla, N. D., Pickands III, J., Smith, R. L. (1989): *[Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone]: Are ozone exceedance rates decreasing? Comments. Rejoinder*. Stat. Sci. **4**^{4}, 378–393. See **smith1989**.