Computer Simulations and the Replication Crisis

Based on the representational capacity of scientific models, critically discuss how computer simulations address the question of replication crisis (also called replicability crisis) in Science and Engineering.

Davey Jackson

Introduction

The replication crisis is one of the biggest issues facing modern science, where many groundbreaking results in a wide range of fields have not been replicated and when replication has been attempted it failed to produce similar results (Ioannidis, 2005). Whilst the only true way to solve this crisis is to replicate studies, that requires significant amounts of investment in time and money. In this essay I will be providing a framework for using models generated from these studies and exploratory techniques in computer simulations to be able to build a metric of uncertainty, absent statistical tests, to quantify which results are more or less likely to be true, which can allow us to prioritise replication resources more effectively or know which unreplicated studies to be more wary of.

The Replication Crisis

The replication crisis is a very complex scientific phenomenon, with myriad causes, mechanisms and effects across a wide range of fields. In this essay our specific focus is going to be on how it has shown our current statistical methods to be insufficient (Ioannidis, 2005). The accepted standard in science is to use a p-value (a measure of how likely a result is to be random chance) of 0.05, meaning a 5% chance that the results were random, as the signifier that a result is significant (Ioannidis, 2005). Even if scientists were purely altruistically looking for the truth, 1 in 20 initial results are probably flukes under such a system, and the pressures of academia and “publish or perish” means that they might engage in p-hacking, where a wide range of measurements are taken and through sheer size of data you can find some statistically significant result (Ioannidis, 2008).

Quantifying Uncertainty

The ideal solution is replication, but if that is not possible then it is necessary to find some way of quantifying uncertainty outside of the normally accepted statistical methods, to classify if a study should be subject to further review or accepted until it can be replicated. Such analysis needs not be restricted to purely the observed data, and in this essay, I will be looking at the underlying theory of the paper and the conclusions drawn from it.

If we cannot use data to classify the uncertainty, we must use some other epistemic source of certainty to compare against, and in this paper, I will be relying on trusting the existing knowledge of a field. One can consider that a scientific field is an aggregate of all the research and analysis that has happened up to this point, and the sum of that analysis is a very certain thing. We can then compare theories to their field, and if they contradict or reject those findings, we can assign the paper a higher uncertainty as it is going against a known high certainty finding.

The specific metrics that I will be arguing for to quantise uncertainty are concreteness, consistency and coherency.

Concreteness is the first metric to look at, as the other two cannot really be evaluated on a non-concrete theory. It refers to the ability of the paper to produce a theory that is well defined and concrete in how it relates the independent variables to the dependent variables. If such a relationship can't be well defined then it means that either we lack the requisite data to understand the relationship, that there is some mechanism that we don't know of or that there is not actually a relationship there. All 3 of those suggest some significant gap in our knowledge that the paper cannot answer, and so a very high uncertainty.

Consistency is a metric that looks at how consistent the results are with the rest of the field. This is both in the direct results the paper produces, and how those compare to similar works, but also when applying the conclusions to broader problems or datasets does the theory align with other works or does it produce results that are unexpected. This is perhaps the simplest metric to justify, as it is a direct comparison to our source of certainty.

Coherency is a metric of the "quality" of the proposed theory. That is to say if a theory produces the kinds of patterns in output and sensitivity to input that is expected for the field. What is expected would obviously depend significantly on the field and the specifics, but generally examples could be: sensitivity to initial conditions, specifically ones that aren't expected to remain constant; odd distributions, where see variance in independent variables produce radical changes that don't follow expected patterns; or unreasonable extrapolations, where going beyond the measured data produces sudden and massive departures from measured data. We can view the expected kinds of patterns of a field as being indicative of the macro structure of the field's subject matter, and those patterns as the current knowledge of how said structure manifests in data and models. If a theory is wildly outside of such patterns, then either the rest of the field is incorrect or massively incomplete or the theory is wrong or extremely exceptional, both of which create uncertainty.

The use of models

It is however impossible to directly test a theory, as an abstract scientific concept, and so this is where models become useful. Models serve as representations of the theories that they are based on, and whilst there is not a 1:1 comparison, by looking at the results of manipulating a model we can draw conclusions about the underlying theory (Contessa, 2007). This is important as models, unlike theories, can be manipulable objects, and so we can vary the parameters of our models to try and view the various behaviours outlined in the previous sections, and if the behaviour of the models match up with the behaviours associated with greater uncertainty, then we can assign the theories greater uncertainty.

Given the requirements to make a manipulable model are the same requirements that underly our concreteness metric, we will only be looking at consistency and coherency, as an unconcrete theory is functionally less certain than any concrete theory even if it has less coherency or consistency.

Computer Simulations

With the theoretical groundwork of uncertainty established and how we can use models to determine the uncertainty of a theory by proxy, we can now discuss the actual testing mechanism, which is computer simulation. Computer simulations are extremely powerful in this regard as we can explore a very wide range of possibilities and states. Even for a fairly simple model, say a theory that produces a simple equation that can be analytically solved, solving the equation by hand would only allow testing a few simple input variables. With a computer simulation we could create detailed plots showing how the results vary across a wide range and multiple independent variables, which can give us an overall output curve that demonstrates the true shape of the model, and therefore equation, and allows us to see if that shape makes sense. The breadth and depth of model analysis allowed by computer simulations is what can allow us to measure the coherency of our model, viewing overall shape, behaviour across a continuous range of inputs and how the model reacts to changes in initial variables.

Computer simulations also allow us to do composite simulations, where we can take an existing simulation based off of accepted theories and their respective models and apply our new model to the simulation as well, to view the results. If we have a strong agent-based simulation of the economy, and a new paper about economic actors is published, we can incorporate their findings into the simulation and run the new composite simulation. If this composite simulation better matches empirically viewed results or the expect outputs then we can be more sure that the model is consistent, and if the model radically departs from expected values then we know that it is likely inconsistent.

Conclusion

In this essay I have shown how traditional statistical conceptions of uncertainty are failing us under the replication crisis, and of the need for another separate measure of uncertainty. I have shown how effective way we can generate this new measure of uncertainty is by comparing these theories to the existing body of knowledge in a field. I have shown a set of specific metrics that can apply that epistemic reasoning, and how by taking theories, building models and then running computer simulations to view the models across a wide range of input values or alongside other models we can measure these metrics and compute uncertainty. In showing that I have provided a strong argument for how computer simulations can be used to effectively help resolve the replication crisis.

Bibliography

Contessa, G. (2007). Representing Reality: The Ontology of Scientific Models and Their Representational Function. PhD. London School of Economics.

Frigg, R. and Nguyen, J. (2021). 'Scientific Representation'. in Zalta, E. N. (ed.) The Stanford Encyclopedia of Philosophy. Winter 2021. Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/win2021/entries/scientific-representation/ (Accessed: 17 April 2024).

Hartmann, S. (2005). 'The World as a Process: Simulations in the Natural and Social Sciences'. in.

Hepburn, B. and Andersen, H. (2021). 'Scientific Method'. in Zalta, E. N. (ed.) The Stanford Encyclopedia of Philosophy. Summer 2021. Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/sum2021/entries/scientific-method/ (Accessed: 17 April 2024).

Ioannidis, J. P. A. (2005). 'Why Most Published Research Findings Are False'. PLoS Medicine, 2 (8), p. e124. doi: 10.1371/journal.pmed.0020124.

Ioannidis, J. P. A. (2008). 'Why Most Discovered True Associations Are Inflated'. Epidemiology, 19 (5), pp. 640-648. doi: 10.1097/EDE.0b013e31818131e7.

Oreskes, N., Shrader-Frechette, K. and Belitz, K. (1994). 'Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences'. Science, 263 (5147), pp. 641-646. doi: 10.1126/science.263.5147.641.

Winsberg, E. (1999). 'Sanctioning Models: The Epistemology of Simulation'. Science in Context, 12 (2), pp. 275-292. doi: 10.1017/S0269889700003422.

Winsberg, E. (2022). 'Computer Simulations in Science'. in Zalta, E. N. and Nodelman, U. (eds) The Stanford Encyclopedia of Philosophy. Winter 2022. Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/win2022/entries/simulations-science/ (Accessed: 17 April 2024).