Software, Computational Complexity, and Randomised Controlled Trials

John Chalisque Allsup

April 25, 2017

1 Introduction

The starting point of this was a Twitter exchange regarding the degree to which Randomised Controlled Trials were sufficiently controlled, and in particular in Mental Health, as compared to other areas of physical health. The other place where concerns such as this may be valid is in genetics. The basic intution is that, when a system is 'computation capable', the degree to which a Randomised Controlled Trial (hereafter RCT) approach can be sufficiently 'controlled' becomes a major issue, and one where it is tempting to simply assume that having a control group is good enough.

The cognitive complexity of the brain is unparalleled in other areas of the human body, and thus there are no proper analogues in human health, that do not themselves heavily involve brain behaviour, to show how computational complexity can wreak havoc with predictability. The point to make is that it is an unavoidable aspect of healthy working brain, and thus these issues can't ignored in the particular case of Mental Health.

The purpose in writing this short article is to illustrate, using computer software, how the 'logic' of reasoning about Mental Health issues from RCTs can, at least in principle, fail spectactularly. The issues about computation which lead to this essentially date to the beginnings of modern computation and the work of mathematicians such as Church and Turing.

2 A Simple Toy Example

Consider a very small computer. It has 1024 bytes of program memory, and 1024 bytes of working memory, and a simple output device capable of printing a string of numbers, each of which is between 0 and 255 inclusive. At the start of the program's running, the working memory is set to be identically zero, and the computer is fully deterministic, so that the behaviour is determined entirely by those 1024 bytes.

Now what happens if we try to do the equivalent of a Randomised Controlled Trial of a number of these machines, which have similar, but *not identical*, contents in their program memories?

If we assume that every configuration of bytes in program memory is, in some way, a valid program, and we are interested in the effect that changing the bytes at locations 100 through 107 (inclusive) has upon the behaviour of the program, then what happens?

In the most general case we have 1024 independent degrees of freedom as to the contents of the program memory, and we have 8 experimental variables, each of which takes a value between 0 and 255 inclusive. There are 256^8 possible values of the 8 experimental variables, and 256^{1024-8} possible values of the remaining bytes in program memory.

Now potentially every single one of thse 1024 bytes matters (since we can read from that area, potentially copy fragments of program memory into working memory, and so on, perform arithmetic upon what we read from program memory, and decide which instruction to execute next based upon the outcome of that arithmetic).

Clearly if we had total knowledge of the 8 experimental variables, but zero knowledge of the rest, we will essentially have no clue as to how the machine, as a whole, will behave. But now consider something analogous to the fact that most PCs run Windows, and a few run Linux. If we take a 'representative sample' of such machines (following the kind of stratification that happens in clinical research), and 'stratify' based upon the contents of program memory at addresses 300 through 303 inclusive, and group based on the number of binary 1's at that location, [1] the statistics of which changes had a 'beneficial effect' and which had a 'negative effect', would present what amounts to an approximation of a weighted average across the disparate classes of configuration (that is, it is like averaging the behaviours of Windows PCs and Linux PCs to get a representative picture of what happens for 'all PCs'). Nobody seriously working in computers thinks along such lines because, on a practical level, to do so is seriously stupid in the absence of deliberately constructed abstraction layers. (And that is why abstract machines like Java were developed.)

The state of modern software engineering shows the current cutting edge of techniques we have to rein in this sort of complexity. Indeed a computer as small as the one in this toy example can be left to a competent programmer to sort out. Once the number of independent degrees of freedom we have to specify the 'configuration space' grows to the scale of even a modern laptop, let alone a human brain, *even if most of the possible configurations are invalid in some way*, we still have a situation where the number of possible 'qualitatively different' configurations is beyond astronomical, and the effect of making changes to some aspect of the configuration depends, *to some variable degree*, upon the particular configuration we are modifying.

Having a control group simply cannot control for this kind of complexity. But this kind of complexity does not usually arise in human beings at all apart from the brain, and

[1] again note that this is a toy example, and it is deliberately silly and unrealistic, so as to make obvious the reasoning issues present

even then, to nowhere near the same degree[2]. My overriding concern about psychiatry as currently practised, and its heavily RCT-based research literature, is that such possibilties haven't been paid proper attention. In a sense, this is akin to what happens if one assumes a *double pendulum* has a behaviour which is simple and measurable as is the case with a *simple pendulum*[3].

3 What about brains and neural nets?

The potential 'logical nightmare' scenario occurs in the situation where *every* change in 'synaptic configuration' of a brain has a qualitatively significant effect upon its behaviour. The extreme opposite is that all brains are broadly equivalent, and that changes amounting to less than 1%, say, of synaptic connections will have a significant change in behaviour.

Here, to keep things conceptually simple, I am assuming that the set of neuroms is the same, and all that varies are the connections and weights, as happens in neural net training in Deep Learning.

The problem you face is that there are significantly more qualitatively different configurations than you have people to sample, and no simple uniformity conditions to rein in the vast and complex space of neural configurations. The cutting edge of Deep Learning have no magic methods here, and generally solve issues by throwing vast amounts of computational resources at them, and observing when the end result is useful. The kind of statistical methods used in Randomised Controlled Trials are of no help here, and the human brain is more complex in many ways than current cutting edge neural nets, *not simpler*.

In the case of neuroscience, which in the Mental Health picture is roughly analogous to the rôle that 'transistor physics and chemistry' plays in electronic computers, things like this latest 'Alzheimers breakthrough' can happen, can be shown in carefully run clinical trials, and get meaningful results. That is largely because degenerative brain disorders create random changes in the neurological structure of the brain and, once the impact of these random changes becomes significant, the mental life of the person is impacted. (I can't help seeing the analogy between what Alzheimers does in its disease process, and what an anti-psychotic, being a dopamine antagonist, does on temporary and changing basis in its claimed therapeutic effect.)

4 A final plea

Please, if you can, explain why such fears are unfounded, or where people have properly thought these things through?

[2]The other potential danger area that comes to mind is DNA. The reason why is the possibility of DNA computing.

[3]A quick read of the Wikipedia page at `https://en.wikipedia.org/wiki/Double_pendulum` should be sufficient to show how different these two are.