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

Do-calculus is a part of probability theory

a comment on Pearl's criticisms [draft]

P.G.L. Porta Mana
Western Norway University of Applied Sciences <pgl@portamana.org>

4 October 2021; updated 7 October 2021

It is shown that the "do-calculus" is already included in probability theory. Some comments about its use.

1 ***

Pearl's "do-calculus" is a very useful and elegant method for specifying statements about causal relationships between physical quantities. Such statements can then be used to draw inferences, from them or about them, given observed data. The do-calculus has been increasingly developed and used in the past three or so decades. Together with its use, however, there has been an increasing misunderstanding about its relation with probability theory, and one frequently reads the claim that *the do-calculus needs, or is, some kind of extension of probability theory*. Statements that point in such a direction appear indeed in Pearl: "The need to adopt a new notation, foreign to the province of probability theory"[1]; "probabilities capture normal relationships in the world, whereas actions represent interventions that perturb those relationships. It is no wonder, then, that actions are treated as foreign entities throughout the literature on probability and statistics; they serve neither as arguments of probability expressions nor as events for conditioning such expressions."[2].

In the present comment I would like to show that such claims are false, and that probability theory does not require any extensions to accommodate the do-calculus. I believe this clarification is important because it concerns the very understanding of what probability theory is about, which is in turn very important for pedagogical reasons.

I believe that this kind of misunderstandings appear because we end up working mechanically with mathematical symbols and objects, forgetting what they mean, "letting our equations lead us by the nose", as Morro wrote to Maxwell[3].

[1] Pearl 2001 § 2.2 p. 23. [2] Pearl 2009 § 4.1 p. 109. [3] Campbell & Garnett 1882 ch. XII, p. 378.

2 The do-calculus within the probability calculus

Imagine that we want to infer the number of cases of a particular disease among individuals belonging to a particular geographic district, at a particular moment in time. Assume that the number of individuals is known, say 1 000, and that the presence or absence of the disease can be assessed with perfect certainty. Our inference is about specific statements, for example 'The number of cases is 500', denoted by N_{500}, and similarly for other numbers.

A clerk gives us the following (reliable) sample values: '10 individuals belonging to the district have the disease, and 10 do not'; denote this statement by V. We ask the clerk how such data were gathered. Now consider two cases:

(i) The clerk tells us '20 individuals were selected, without replacement, from the full register of inhabitants by using a pseudo-random number generator, and then checked for the disease'. Denote this statement by U (for *U*nsystematic selection).

(ii) The clerk says 'My plan was to *select* 10 individuals with the disease, and 10 without'. Denote this statement by C (for Chosen individuals).

The inferences we can make about the incidence of the disease in the district are obviously very different in the two cases. The data we have in the second case are almost uninformative – almost but not completely: we can still deduce that the number of cases is strictly between 10 and 990, rather than between 0 and 1 000. The difference between the two cases is a matter of *intervention*: in the second case the values of the quantities in our data were effectively selected by the clerk.

Probability theory can be used in either case, and it generally leads to different probabilities in the two cases. For example, assuming uniform prior beliefs in the number of cases, and assuming the clerk had inexpensive resources to choose individuals – denote these assumptions by I, we find

$$\text{case (i):} \qquad \mathrm{P}(N_{500} \mid V \land U \land I) = 0.373\% \tag{1}$$

$$\text{case (ii):} \qquad \mathrm{P}(N_{500} \mid V \land C \land I) = 0.102\% \tag{2}$$

That the probabilities must generally be different is clear from the fact that the conditionals in the two cases are different: $V \land U \land I$ vs $V \land C \land I$.

The difference in the statements U and C is very important, and leads also to different probability for the sample values given the hypothesis, for example

$$\text{case (i):} \qquad \mathrm{P}(V \mid N_{500} \land U \land I) = 17.8\% \tag{3}$$

$$\text{case (ii):} \qquad \mathrm{P}(V \mid N_{500} \land C \land I) = 100\% \tag{4}$$

Laplace? Bernoulli, Johnson, Keynes, Jeffreys, Cox, Paris, Van Horn Gaifman, Scott & Kraus, Pólya, Jaynes, Tribus, Hailperin, Maxwell, Adams, Howson

Bibliography

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

Campbell, L., Garnett, W. (1882): *The Life of James Clerk Maxwell: With a Selection from His Correspondence and Occasional Writings and a Sketch of His Contributions to Science*. (Macmillan, London). https://archive.org/details/lifeofjamesclerk00campuoft .

Corfield, D., Williamson, J., eds. (2001): *Foundations of Bayesianism*. (Springer, Dordrecht). doi:10.1007/978-94-017-1586-7.

Pearl, J. (2001): *Bayesianism and causality, or, why i am only a half-Bayesian*. In: Corfield, Williamson (2001): 19–36. doi:10.1007/978-94-017-1586-7_2.

— (2009): *Causality: Models, Reasoning, and Inference*, 2nd ed. (Cambridge University Press, Cambridge). First publ. 2000.