Statistics and Causal Inference by Paul W. Holland

Simon Oergaard, Sascha Valentin Brown, Emilie Jessen, and Luc Voorend
Niels Bohr Institute - Advanced Methods in Applied Statistics
(Dated: March 6, 2024)

INTRODUCTION

We have all heard the saying 'Correlation does not imply causation.' Nevertheless, in our daily discourse, we frequently attribute causes and effects, such as asserting that 'rain causes Julia to open her umbrella.' Some daring scientists may even assert, 'Given that it is raining, and Julia typically opens her umbrella when it rains, it is highly probable that the rain causes Julia to open her umbrella.' However, within the realm of statistics, there has long been a distinct reluctance to draw causal inferences. In his 1986 article, Paul W. Holland endeavors to introduce causality into statistics and the underlying mathematical framework. This exposition focuses on the theoretical framework of the Rubin Causal Model, a model of causal inference statistics, how it diverges from associational inference statistics, and its contemporary applications.

Associational Inference

In associational inference statistics, analysis revolves around a population denoted as U, composed of individual units labeled as u_i. Within this framework, associational variables manifest as functions of $Y(U = u_i)$ and $A(U = u_i)$. These variables serve as the basis for constructing joint distributions $Pr(Y = y, A = a)$ and conditional distributions $Pr(Y = y|A = a)$. The essence of associational inference lies in harnessing these distributions to formulate estimations, conduct tests, derive posteriors, and ultimately infer relationships between variables Y and A across the population U.

Notice that temporal sequencing holds minimal significance within this context. While each u_i entails measurements of $A(U = u_i)$ and $Y(U = u_i)$, there exists no inherent correlation dictating the order of occurrence between the two measurements.

RUBIN CAUSAL INFERENCE

In a manner akin to the aforementioned discussion, our analysis continues within the confines of a population denoted as U, comprising individual units denoted as u_i. We now pivot our attention to two variables, denoted as S and Y_s. Here, S represents the causes or treatments, while Y_s signifies the post-exposure outcomes following treatment S. This discourse delves into a subset of Rubin Causal Inference, specifically addressing two causes denoted as $S = t$ and $S = c$ (representing treatment and control, respectively), with the acknowledgment that further causes can be integrated.

Notably, $Y_t(U)$ quantifies the treatment outcomes across the entire population, while $Y_c(U)$ measures the control outcomes within the same population. This formulation facilitates the conceptualization of the causal effect of treatment t relative to control c as $Y_t - Y_c$.

However, a challenge arises in concurrently measuring Y_c and Y_t. When a unit is affected by cause c and Y_c is measured, one is compelled to pose the counterfactual question: "What would the outcome have been if the same unit had been subjected to treatment t, and Y_t measured?"

This leads to *the fundamental problem of Causal Inference*, asserting the impossibility of observing the values of $Y_t(u)$ and $Y_c(u)$ for the same unit u, simultaneously. Nonetheless, the absence of direct observation does not imply a dearth of knowledge concerning these values. Subsequently, the remainder of this exposition is aimed at solving this quandary, using both scientific- and statistical methods.

Scientific Solutions

The scientific solution to the fundamental problem exploits some assumptions about homogeneity and/or invariance to determine both Y_t and Y_c from experiments.

1. Temporal Stability and Casual Transience assumptions.

Temporal stability is the assumption that the measurement does not change in the interval between measuring $Y_{time=0}(S) = Y_{time=\Delta t}(S)$, while Casual transience is that first testing one cause and then a second does not affect the second. That is, measuring c and then t does not affect the measurement of t, or v.v.

2. Unit Homogeneity.

Unit homogeneity is an assumption which states $Y_c(u_1) = Y_c(u_2)$ and $Y_t(u_1) = Y_t(u_2)$, that is exposing u_1 and u_2 to c will lead to the same measurement. Using these assumptions makes it possible to determine both factual and counterfactual values of Y. With these, causal claims can be made.

Importantly, even though a scientist might be able to convince himself and others that these assumptions are met, the author emphasizes that this is never testable.

Statistical Solution

The statistical solution uses averages of causal effects, which is defined as

$$T = E(Y_t) - E(Y_c)$$

Here, T is the expected causal effect of t relative to c over the population U. This is a powerful tool, since, if some units are exposed to t, they will grant you information about $E(Y_t)$ and v.v. about c. The paper lists two ways of doing this.

1. Independence Assumption

To obtain $E(Y_t)$ one makes the assumption that $E(Y_t) = E(Y_t|S = t)$. The expectation value, denoted as $E(Y_t)$, remains invariant whether computed across the entire population U or exclusively over the subset u that has undergone the specific cause.

To attain this, the experiment has to have a large enough population and the population has to be randomized. This is called the *prima facie causal effect*

$$T_{PF} = E(Y_t|S = t) - E(Y_c|S = c)$$

The important point here is that the statistical solution answers the fundamental problem by estimating the average effect of t over a population of U.

2. Constant Effect

The constant effect assumption is the assumption that all individuals are impacted with the same effect given a cause. $T = Y_t(u) - Y_c(u)$, $\forall u \in U$. Which then leads $T_{pf} = T + E(Y_t|S = t) - E(Y_c|S = c)$. The constant effect then works by mapping $Y_t(u_i) \rightarrow Y_c(u_i)$, to test the unknown effect of the course.

Difference between the Rubin model and associational inference

The key difference between the Rubin model and associational inference is that the latter works with a population that has measurable variables which are attributes of the population. Associational inference makes claims as results of a snapshot in time of the population. An example could be: What is the probability that I am a man, given I study physics? Here the fact that I study physics is treated as an attribute.

In causal inference, we progress from making probabilistic claims based on attributes to determining the effect of a cause. Thus the cause is essential. The variable Y_s can only be measured after a cause S has been imposed on a unit u. This temporal aspect introduces a nuanced complexity, as the sequence of measurement and causation becomes crucial.

WHAT CAN BE A CAUSE

But what is even a cause? Can you actually blame the big bang for being late to work? To answer this, Paul W. Holland sets up 3 examples.
(A) She did well on the exam because she is a woman.
(B) She did well on the exam because she studied hard .
(C) She did well on the exam because she was coached by her teacher.
These examples point out the limitations of the model. In theory, a unit needs to be exposable to both $S = c$ and $S = t$, for c and t to be causes. In general, this limitation has the consequence that attributes cannot never be causes, therefore A is not a cause. In B, it is not clear what is mean by studied hard, and a control can not be defined. As a result, there is no causal relation in examples A and B. Only the cause in example C is twofold (S = Coached by teacher or S = Not Coached by teacher).

OUTLOOK

This paper, authored in 1986, stands as one of the pioneering works in causal inference statistics. Since its publication, considerable advancements have been made in causal modeling, revealing certain weaknesses inherent in the model delineated within this paper.

In the statistical approach featuring a constant causal effect, the determination of the causal effect was achieved by establishing a mapping from $E(Y_t)$ to $E(Y_c)$. Presently, leveraging diverse assumptions, it has become feasible to establish a mapping from $E(Y_t)$ to $E(Y_c)$ without necessitating a constant causal effect. One such avenue involves employing machine learning algorithms with sufficiently extensive datasets.

A flaw in the Rubin Causal Model lies in its inability to compare different causal chains and account for direct and indirect effects [3]. This brings us to Judea Pearl, who over the last 30 years has invented a new branch of mathematics called *Do Calculus* which has rectified the problem of comparing causal chains. One of the goals of Pearl's ongoing research is to make artificial intelligence that can distinguish between cause and correlation. Causal inference statistics is starting to inform areas such as policy decisions, court verdicts etc. and will most likely play a larger role in the future.
However, extracting causal relationships solely from data remains an elusive feat, and the requisite steps to surmount this challenge remain ambiguous [2]. The science of causal inference has gotten far in evaluating the effect of a cause but is still inherently bound to the ominous god telling you how the cause came to be.

Write-up:
- Overall it was a good write-up. While being not very rigorous on technical details, it was clear and there were very few typos and editing errors.
Presentation:
- In the Rubin model is it necessary that 'c' is a non-treatment? If 't' and 'c' are both treatments is it possible to have causal inference for one of t or c, i.e. is Y_t - Y_c requiring that one is not an action?
The presentation example of radiation (which I thought was good and easily relatable) was good and the assumptions (temporal stability, causal transience, and homogeneity) was well covered and appropriate.
The presentation could have been more technical in the mathematical formulism with an example too.
Nice job responding to questions about; especially about the assumptions. .
Grade: 5/5 write-up and presentation 4.5/5

[1] Paul W. Holland *Statistical and Causal inference* (Journal of the American Statistical Association, Vol. 81, No.396 (Dec., 1986), pp. 945-960 https://www.jstor.org/stable/2289064)

[2] Judea Pearl *Causal inference in statistics: An overview* (Computer Science Department University of California, Los Angeles, September 2009)

[3] Judea Pearl *Direct and Indirect Effects* (Cognitive Systems Laboratory Computer Science Department, June 2001)