Probabilistic Variational Causal Effect

Usef Faghihi & Amir Saki

Département de Mathématiques et d'Informatique
Université du Québec à Trois-Rivières

November 2024

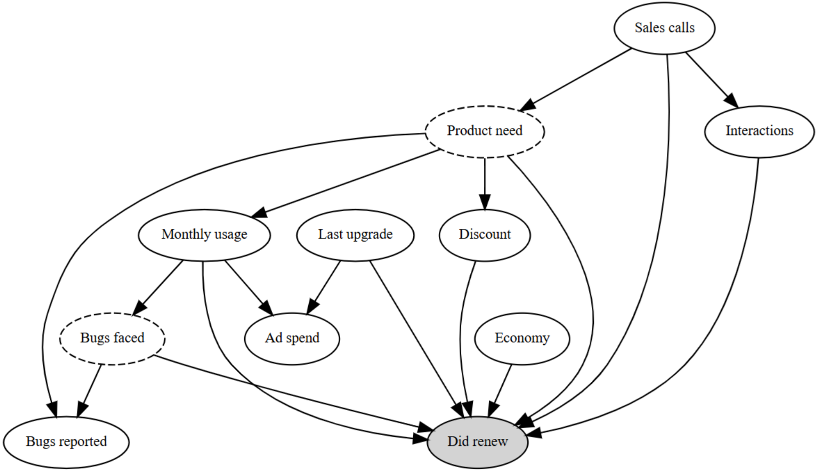
Université du Québec à Trois-Rivières

Table of Content

In this presentation we are looking for a causal inference metric that uses one mathematical metric that accounts for:

- Consideration of various types of treatment variables, such as binary, categorical, and continuous.
- Accounting not only for interventions in treatment values but also for their rarity and frequency.
- Controlling the rarity and frequency of treatments by introducing a degree d.
- Controlling the rarity and frequency of covariates by introducing a degree r.
- Gradient: captures how the outcome responds to infinitesimal interventions on the variables of interest. Therefore, since these outcome changes result from deliberate interventions, they are causal.

Did Renew Dataset



Lundberg, Scott, E. Dillon, J. LaRiviere, J. Roth, and V. Syrgkanis. "Be careful when interpreting predictive models in search of causal insights." Towards Data Sci (2021): 1-15.

Table of Content

"Causality is the science of cause and effect, focusing on how to interpret and quantify the impact of one variable on another when one actively intervenes. This means understanding not only that two variables are related, but also whether and how changing one variable will produce changes in another."

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.

We want to find the causal effect of X on Y. Assume $Y = g(X, Z)$.

Table of Content

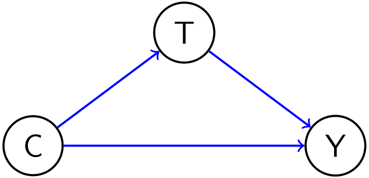
We start with an example. Suppose a researcher wants to know if a specific pill is effective against lung cancer. The *causal effect* of the pill on cancer could be calculated as the difference between the patient's state in the following cases: taking the pill and not taking it.

$$\text{Causal Effect} = \mathbb{E}[Y|\text{do}(\text{Pill}) = 1] - \mathbb{E}[Y|\text{do}(\text{Pill}) = 0]$$

Average Treatment Effect (ATE)

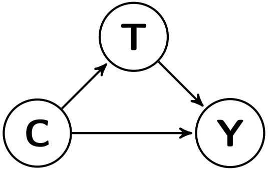
The ATE aggregates the treatment effects across all subgroups c, weighted by their probabilities:

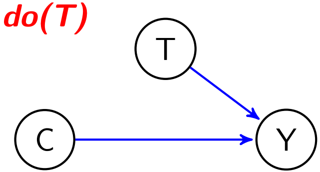
$$\text{ATE} = \sum_c [\text{Proportion in subgroup } c \times \text{Treatment effect in subgroup } c]$$



$$p(y|t) = \frac{1}{P(t)} \sum_c P(C)P(t|c)P(y|t,c)$$

Thus, for a DAG like the one below, we have the following formula:





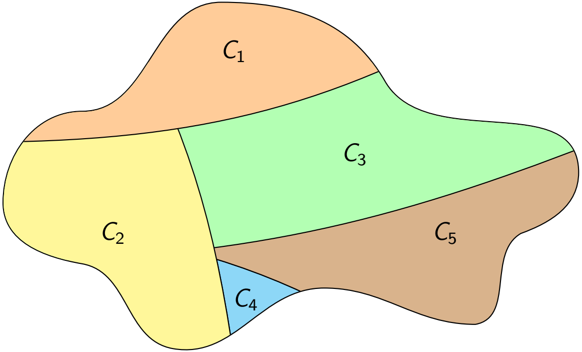
$p(y|do(t)) = \sum_c P(C)P(y|t, C)$

$$\mathbb{P}(y|do(t_0)) = \sum_c \mathbb{P}(c)\mathbb{P}(y|c, t_0),$$

$$\mathbb{E}(Y|do(t_0)) = \sum_y y\mathbb{P}(y|do(t_0)) = \sum_y y \left(\sum_c \mathbb{P}(c)\mathbb{P}(y|c, t_0) \right)$$

$$= \sum_c \mathbb{P}(c) \left(\sum_y y\mathbb{P}(y|c, t_0) \right) = \sum_c \mathbb{P}(c)\mathbb{E}(Y|c, t_0)$$

Table of Content

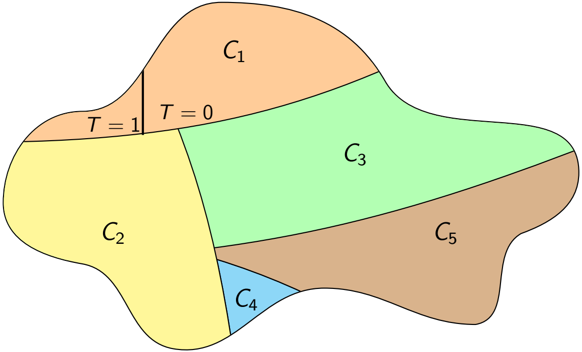


$$ATE = \sum_c \mathbb{P}(c)\left(\mathbb{E}(Y|c,1) - \mathbb{E}(Y|c,0)\right)$$

If we have the following situation

$\mathbb{P}(c_4)$ very small,
$\mathbb{E}(Y|c_4,1) - \mathbb{E}(Y|c_4,0)$ very very large,
$\mathbb{E}(Y|c,1) - \mathbb{E}(Y|c,0)$ very very small,
$c = c_1, c_2, c_3, c_5$

Then, the ATE is large! So, one might ask: Is this small part $C = c_4$ really so important that it can change the value of the ATE so much?



- Degree r: Controls the rarity and frequency of each subpopulation C_1 (normalized subpopulation determined by c).
- Degree d: Controls the rarity or frequency of receiving treatment within each subpopulation.

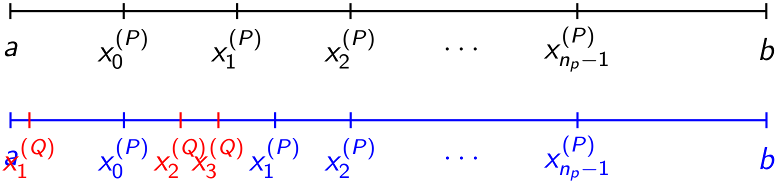
The Total Variation of a Univariate Function

Definition:

Let $f : [a, b] \to \mathbb{R}$ be a real-valued function defined on a closed interval $[a, b]$. A partition P of the interval $[a, b]$ is an ordered set of points:

$$P = \{a = x_0^{(P)} < x_1^{(P)} < \cdots < x_{n_P-1}^{(P)} < x_{n_P}^{(P)} = b\}$$

The set of all possible partitions of $[a, b]$ is denoted by $\mathcal{P}([a, b])$.



The total variation $V(f)$ of the function f is defined as:

$$V(f) := \sup_{P \in \mathcal{P}([a,b])} \sum_{i=1}^{n_P} \left| f\left(x_i^{(P)}\right) - f\left(x_{i-1}^{(P)}\right) \right|$$

This means we consider all possible partitions P of the interval $[a, b]$, compute the sum of absolute differences $\left| f\left(x_i^{(P)}\right) - f\left(x_{i-1}^{(P)}\right) \right|$ for each partition, and take the supremum (the least upper bound) of these sums.

Properties:

If f is continuously differentiable on $[a, b]$, then the total variation can be calculated as:

$$V(f) = \int_a^b |f'(t)| \, dt$$

This integral represents the arc length of f when viewed as a curve in one-dimensional space.

Interpretation:

- The total variation $V(f)$ measures the total amount by which the function f increases or decreases over the interval $[a, b]$.
- It can be thought of as the total "distance" traveled by the function's values as t moves from a to b.
- In physical terms, if $f(t)$ represents the position of a particle moving along a straight line over time t, then $V(f)$ is the total distance the particle travels between times a and b.

Total Variation in the Discrete Case

Definition:

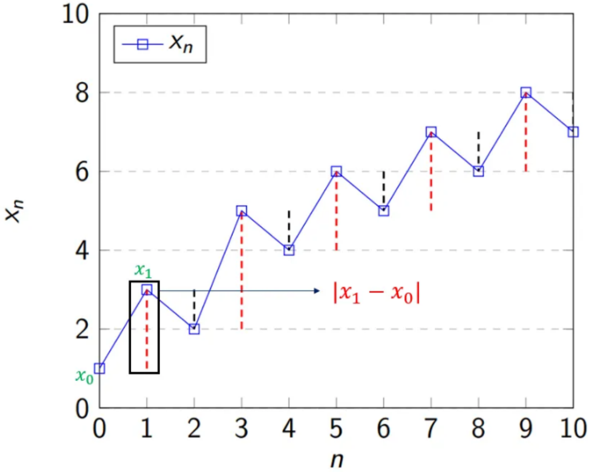
In the discrete setting, consider a sequence of real numbers $x_1, x_2, \ldots, x_n$.
The total variation TV of this sequence is defined as:

$$\text{TV} = \sum_{i=1}^{n-1} |x_{i+1} - x_i|$$

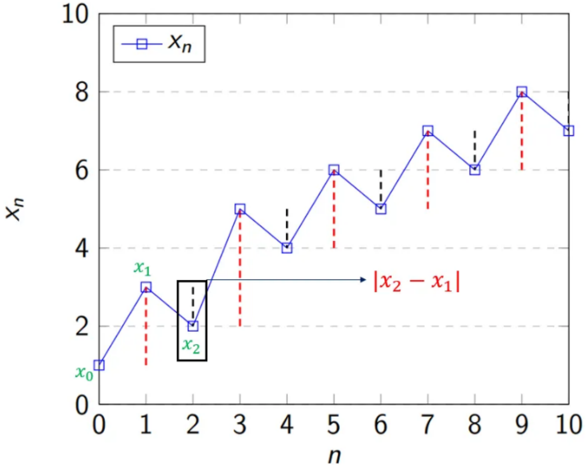
Interpretation:

- Total variation in the discrete case measures the cumulative absolute difference between consecutive elements in the sequence.
- It quantifies how much the sequence "varies" or "fluctuates" as you move from one term to the next.

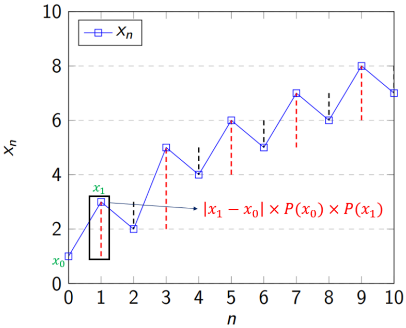
Total Variation of a Sequence



Total Variation of a Sequence



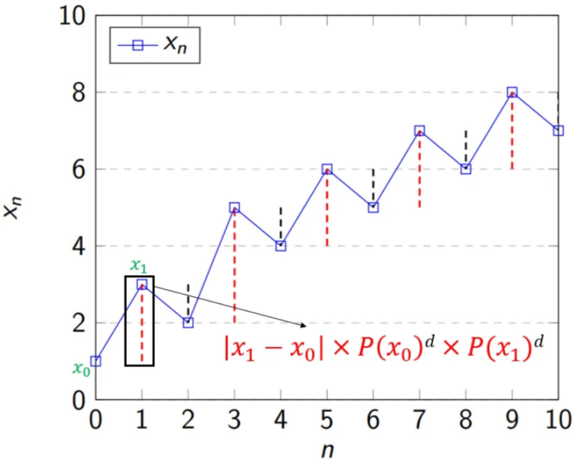
Probabilistic Total Variation



We select x_0 andx_1 independently as an intervention

Table of Content

Total Variation of degree d



PACE of degree d

For a finite random variable x.
The expected value of the following formula with respect to $\mathbf{Z}$ is called the PACE of degree d :

$$\mathcal{PIV}^{\mathbf{z}}(X \to Y) := \max_{P \in \mathcal{P}} \sum_{i=1}^{n_P} \left| g_{in}(x_i^{(P)}, \mathbf{z}) - g_{in}(x_{i-1}^{(P)}, \mathbf{z}) \right| \mathbb{P}(x_i^{(P)} | \mathbf{z})^d \mathbb{P}(x_{i-1}^{(P)} | \mathbf{z})^d.$$

- g_{in} is g when we intervene on the value of x.
- $\mathcal{P}$ the set of all ordered subsets of values of X.

Variations totales positive et négative

Now, we briefly explain an intuition behind the degree d in our PACE metric. Let us assume that $d_2 > d_1 > 0$. Also, assume that $1 \leq i, j \leq n_p$ are in such a way that

$$\left(\frac{P(x_{i-1} \mid z)P(x_i \mid z)}{P(x_{j-1} \mid z)P(x_j \mid z)}\right)^{d_1} > 1.$$

In the nominator, we have the availability of changes from x_{i-1} to x_i, degree of d_1. In the denominator, we have the availability of changes from x_{j-1} to x_j, degree of d_1. That is the availability of changes from x_{i-1} to x_i is higher than availability of changes from x_{j-1} to x_j.

Variations totales positive et négative

Then, we have that

$$\left(\frac{P(x_{i-1}^{(P)} \mid z)P(x_i^{(P)} \mid z)}{P(x_{j-1}^{(P)} \mid z)P(x_j^{(P)} \mid z)}\right)^{d_2} > \left(\frac{P(x_{i-1}^{(P)} \mid z)P(x_i^{(P)} \mid z)}{P(x_{j-1} \mid z)P(x_j \mid z)}\right)^{d_1}$$

if and only if

$$\left(\frac{P(x_{i-1}^{(P)} \mid z)P(x_i^{(P)} \mid z)}{P(x_{j-1}^{(P)} \mid z)P(x_j^{(P)} \mid z)}\right)^{d_2-d_1} > 1,$$

If we have $(\alpha^d) > 1$, then α also is > 1. Thus in the above, since

$$\left(\frac{P(x_{i-1}^{(P)} \mid z)P(x_i^{(P)} \mid z)}{P(x_{j-1}^{(P)} \mid z)P(x_j^{(P)} \mid z)}\right) > 1,$$

then to any positive power, the division is > 1. So when we make changes, the rate of changes from x_{i-1} to x_i compared to x_{j-1} to x_j increases, which means that the more frequent events have a bigger causal impact.

why do we used d to the power

In order to observe the casual effect changes, we are increasing the importance of the frequent events exponentially.

Other possibilities degree d

When we incorporate the degree d as the exponent of the probability values, we are exponentially increasing the importance of more frequent treatment values. We can also use logarithm function.
More general than the previous generalization, One could further adjust this importance by using other functions instead of exponential functions. Thus, we can generalize the definition to have the following terms:

$$|g_{in}(x,z) - g_{in}(x',z)|\, w(x_0, x_1, d),$$

where $w(x_0, x_1, d)$ is a weight function that allows for more manual control over the importance of treatment value changes.

Table of Content

Controlling the rarity or frequency of sub-populations

To control the importance of the rarity or frequency of subpopulations, we introduce a new degree r, in addition to the previously considered degree d. Specifically, we replace the probability values $\mathbb{P}(c_0), \dots, \mathbb{P}(c_k)$ of observing all possible values of C with $\mathbb{P}(c_0)^r/A, \dots, \mathbb{P}(c_k)^r/A$, where

$$A = \mathbb{P}(c_0)^r + \cdots + \mathbb{P}(c_k)^r.$$

Thus, we define the generalized PACE of degree (d, r) as follows:

$$\mathrm{PACE}_d^{(r)}(X \to Y) := \mathbb{E}_{c \sim \mathbb{P}_C^r/A}\left[\mathcal{PIV}_d^c(X \to Y)\right].$$

Controlling the rarity or frequency of sub-populations

$$\mathbb{E}_{c\sim\mathbb{P}^r_C/A}\left[\mathcal{PIV}^c_d(X\to Y)\right]$$

$$=\sum_c\left[\mathcal{PIV}^c_d(X\to Y)\right]\left(\frac{p^r(c)}{\sum_i p^r(c_i)}\right)$$

- $\mathbb{E}_{c\sim\mathbb{P}^r_C/A}[\cdot]$:
  which means we are taking the expected value (average) of a function over the variable c, where c is sampled according to the adjusted probability distribution $\mathbb{P}^r_C/A$.

This approach ensures that our causal inferences are not unduly influenced by the distribution of sub-populations, allowing for more nuanced and equitable analyses.

Total Positive and Negative Variations

Let $r \in \mathbb{R}$. Define $r^+ := \max\{r, 0\}$ and $r^- := |r| - r^+$. Then, the total positive variation of a function $f : [a, b] \to \mathbb{R}$ is defined as follows (see [**9**]):

$$\mathcal{V}(f)^+ := \sup_{P \in \mathcal{P}([a,b])} \sum_{i=1}^{n_P} \left(f(x_i^{(P)}) - f(x_{i-1}^{(P)}) \right)^+ .$$

Similarly, the total negative variation of f is defined as follows:

$$\mathcal{V}(f)^- := \sup_{P \in \mathcal{P}([a,b])} \sum_{i=1}^{n_P} \left(f(x_i^{(P)}) - f(x_{i-1}^{(P)}) \right)^- .$$

We can define $\overline{\mathbb{PEACE}}_d(X \to Y)$, as the mean of the following MEAN VARIATION:

$$\overline{\mathcal{PIEV}}^{\mathbf{z}}_d(X \to Y) := \sum_{i=1}^{l} |g_{in}(x_i, \mathbf{z}) - g_{in}(x_{i-1}, \mathbf{z})| \, \mathbb{P}_d(x_{i-1}, x_i \mid \mathbf{z})$$

where

$$\mathbb{P}_d(x_{i-1}, x_i \mid \mathbf{z}) := \frac{\mathbb{P}(x_i \mid \mathbf{z})^d \mathbb{P}(x_{i-1} \mid \mathbf{z})^d}{\sum_{j=1}^{l} \mathbb{P}(x_j \mid \mathbf{z})^d \mathbb{P}(x_{j-1} \mid \mathbf{z})^d}, \quad i = 1, \ldots, l.$$

It normalizes the transition probabilities across all x_i, x_{i-1} pairs.
Similarly, we can define the MEAN PACE as well as the positive and the negative versions of all of these MEAN metrics.

PEACE and Mean PEACE

PEACE (Probabilistic Easy Variational Causal Effect): This metric assesses the causal impact of an intervention in terms of the total variations in outcomes due to changes in a treatment variable X, while keeping the covariates Z constant. PEACE considers each incremental change in the treatment variable along a specific sequence (e.g., from x_0 to x_1, x_1 to x_2, and so on), weighted by the probability of these treatment values under a particular covariate condition. Thus, PEACE measures the overall effect across the range of treatment changes available in a given setting, accounting for each possible transition.

Mean PEACE: Can provide insights into how the overall causal effect might behave in the average scenario, which is particularly useful in cases where some transitions in treatment values are more frequent or probable than others.

PACE vs. PEACE

PACE: PACE considers the **highest possible** interventional changes by examining all partitions of the values of X and selecting the maximum interventional effect across these partitions. This approach captures a broader spectrum of potential variations and emphasizes the maximum observed change in Y as X varies.

PEACE: PEACE, on the other hand, focuses on **smaller, incremental changes** in X, aggregating the causal changes in Y by taking a weighted sum of these incremental effects. This method is simpler as it uses a fixed partition for X, calculating the causal effect through the summation of smaller variations without selecting a maximum.

Other Formulas

$$\mathcal{PIEV}_d^{\mathbf{z}}(X \to Y) = \sum_{i=0}^{l} |g_{in}(x_i, \mathbf{z}) - g_{in}(x_{i-1}, \mathbf{z})| \mathbb{P}(x_i|\mathbf{z})^d \mathbb{P}(x_{i-1}|\mathbf{z})^d$$

$$\mathbb{PEACE}_d(X \to Y) = \mathbb{E}_{\mathbf{Z}} \left(\mathcal{PIEV}_d^{\mathbf{z}}(X \to Y) \right)$$

$$\mathcal{PIV}_d^{\mathbf{z}}(X \to Y) = \sup_{P \in \mathcal{P}} \sum_{i=0}^{n_P} |g_{in}(x_i^{(P)}, \mathbf{z}) - g_{in}(x_{i-1}^{(P)}, \mathbf{z})| \mathbb{P}(x_i^{(P)}|\mathbf{z})^d \mathbb{P}(x_{i-1}^{(P)}|\mathbf{z})^d$$

$$\mathbb{PACE}_d(X \to Y) = \mathbb{E}_{\mathbf{Z}} \left(\mathcal{PIV}_d^{\mathbf{z}}(X \to Y) \right)$$

$$\mathcal{SPIV}^{\mathbf{z}}(X \to Y) = \sup_{x < x'} |g_{in}(x, \mathbf{z}) - g_{in}(x', \mathbf{z})| \mathbb{P}(x|\mathbf{z})^d \mathbb{P}(x'|\mathbf{z})^d$$

$$\mathbb{SPACE}_d(X \to Y) = \mathbb{E}_{\mathbf{Z}} \left(\mathcal{SPIV}_d^{\mathbf{z}}(X \to Y) \right)$$

$$\mathcal{APIV}^{\mathbf{z}}(X \to Y) = \sum_{x' < x} |g_{in}(x, \mathbf{z}) - g_{in}(x', \mathbf{z})| \mathbb{P}(x|\mathbf{z})^d \mathbb{P}(x'|\mathbf{z})^d$$

$$\mathbb{APACE}_d(X \to Y) = \mathbb{E}_{\mathbf{Z}} \left(\mathcal{APIV}_d^{\mathbf{z}}(X \to Y) \right).$$

Differences between Versions of PACE Metrics

In the provided formulas, different versions of PACE metrics assess variations in causal effects between a treatment variable X and outcome Y under different approaches. Here's an explanation of the differences:

- **PEACE (Probabilistic Easy Variational Causal Effect)**: Focuses on assessing incremental causal effects across smaller changes in X values, taking a weighted sum of these changes while keeping covariates Z constant. PEACE considers smaller variations, aggregating their impact in a straightforward manner without focusing on maximal changes.
- **PACE (Probabilistic Average Causal Effect)**: Generalizes PEACE by selecting maximal interventional effects across possible partitions of X. This approach captures more pronounced changes in Y by considering all partitions and focusing on the maximum effect observed. It provides a broader spectrum of potential variations compared to PEACE.
- **SPIV and SPACE (Supremum Probabilistic Interventional Variation and Supremum PACE)**: Capture the maximum interventional effects across single, specific transitions in X. SPACE aggregates SPIV effects across different values of Z, emphasizing the highest interventional changes rather than average or smaller increments.
- **APIV and APACE (Aggregated Probabilistic Interventional Variation and Aggregated PACE)**: Further extend the concepts by considering all possible transitions between X values, rather than specific or incremental ones. This comprehensive approach aggregates interventional effects across every potential change in X values, providing a holistic view of causal impacts.

In summary:

- **PEACE** focuses on incremental, weighted sums.
- **PACE** generalizes to consider maximal changes across partitions.
- **SPIV/SPACE** focus on the maximum changes for specific transitions.
- **APIV/APACE** aggregate effects across all possible transitions, offering the most comprehensive view.

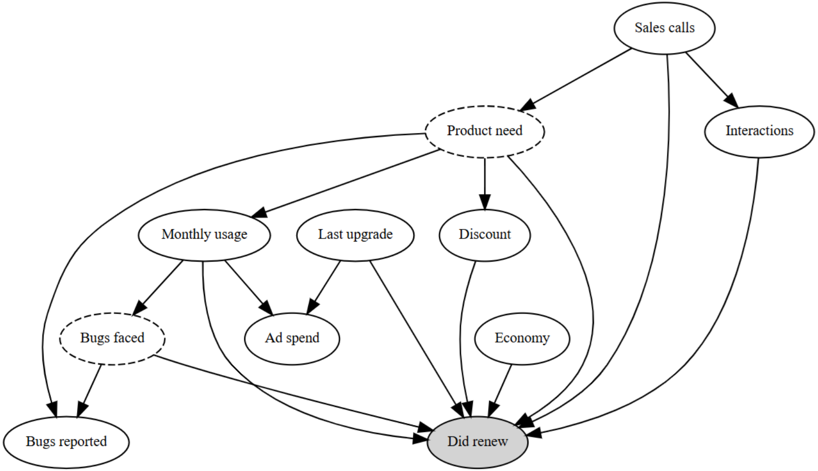
Understanding Different Versions of PACE Metrics

- **PEACE (Probabilistic Easy Variational Causal Effect):**
  - Example: Imagine you're studying the effect of daily coffee consumption (cups per day) on energy level. By measuring small changes (1 cup to 2 cups, etc.) and keeping other factors constant, PEACE assesses the *total impact* of each incremental change, weighted by the probability of each level (since you might rarely drink 5+ cups but often drink 1-3 cups).
  - PEACE measures the sum of small effects, capturing the cumulative impact of all incremental changes.
- **PACE (Probabilistic Average Causal Effect):**
  - Example: Studying the effect of exercise frequency on body weight. By analyzing categories like "no exercise," "light," "moderate," and "intense," PACE calculates the average effect on weight while emphasizing the **highest possible interventional effect**.
  - This approach captures major shifts, like the difference in body weight between "no exercise" and "intense exercise."

Understanding Different Versions of PACE Metrics

- **SPIV (Supremum Probabilistic Interventional Variation) and SPACE (Supremum PACE):**
  - Example: Assessing how different levels of sugar intake affect blood glucose. SPIV identifies the *largest single interventional change* (e.g., the largest increase in blood glucose between specific sugar levels).
  - SPACE aggregates SPIV values across scenarios (e.g., sedentary vs. active), providing an overall view of the highest interventional changes across settings.
- **APIV (Aggregated Probabilistic Interventional Variation) and APACE (Aggregated PACE):**
  - Example: Studying how increasing doses of medication impact symptom relief. APIV sums the effects of all dose changes (e.g., from 10 mg to 20 mg, etc.).
  - APACE aggregates these effects across groups (e.g., adults vs. seniors), offering a comprehensive view of overall causal effects across all changes and subgroups.

Did Renew Dataset



Lundberg, Scott, E. Dillon, J. LaRiviere, J. Roth, and V. Syrgkanis. "Be careful when interpreting predictive models in search of causal insights." Towards Data Sci (2021): 1-15.

Code on GitHub

The code for this project is available on GitHub. Click the link below to access it:

GitHub Repository

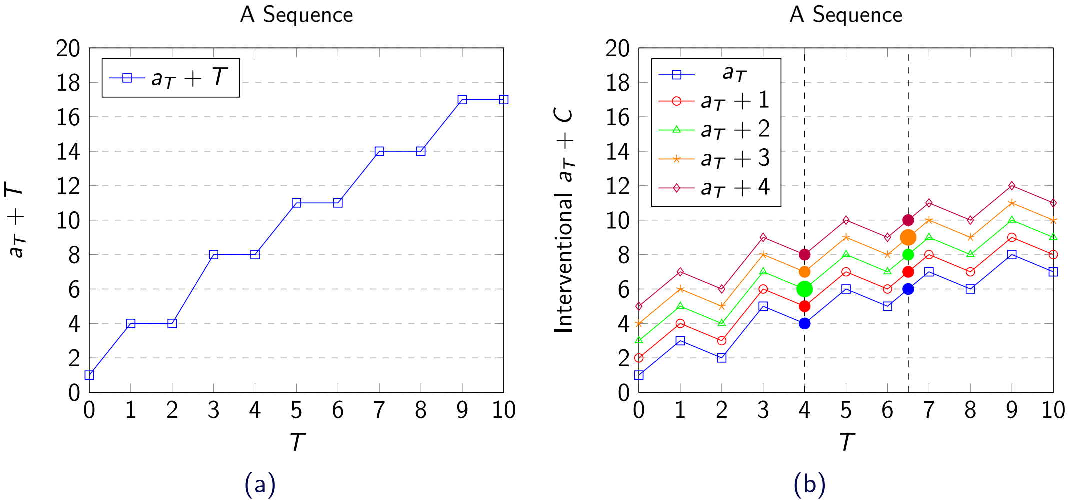
References

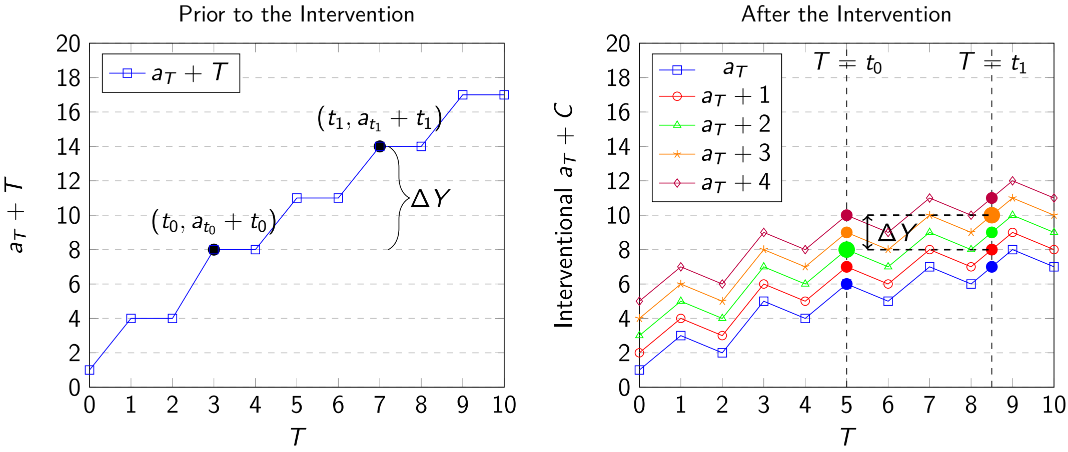
[1] G. W. Imbens, D. B. Rubin, Causal inference in statistics, social, and biomedical sciences, Cambridge University Press, 2015.
[2] D. B. Rubin, Estimating causal effects of treatments in randomized and nonrandomized studies., Journal of educational Psychology 66 (5) (1974)688. [3] J. Pearl, D. Mackenzie, The book of why: the new science of cause and effect, Basic books, 2018.
[4] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, B. Sch¨olkopf, Quantifying causal influences, The Annals of Statistics 41 (5) (2013) 2324–2358.
[5] C. Jordan, Sur la s´erie de fourier, Comptes rendus hebdomadaires dess´eances de l'Acad´emie des sciences (in French) 92 (1881) 228–230.
[6] N. Dunford, J. T. Schwartz, Linear operators. I. General theory. (Withthe assistence of William G. Bade and Robert G. Bartle), Pure and Applied Mathematics. Vol. 7. New York and London: Interscience Publishers.xiv, 858 p. (1958). (1958).

References

[7] S.Wright, The relative importance of heredity and environment in determiningthe piebald pattern of guinea-pigs, Proceedings of the National Academy of Sciences 6 (6) (1920) 320–332.
[8] S. Wright, Correlation and causation (1921).
[9] K. A. Bollen, J. Pearl, Eight myths about causality and structural equation models, in: Handbook of causal analysis for social research,Springer, 2013, pp. 301–328.
[10] S. Boslaugh, "Structural Equation Modeling". Encyclopedia of epidemiology, Sage Publications, 2007.
[11] O. D. Duncan, Introduction to structural equation models, Elsevier, 2014.
[12] F. W. English, "Structural Equation Modeling". Encyclopedia of educational leadership and administration, Sage publications, 2006.

Thank you very much for your time and attention!


A Sequence
$a_T + T$
$a_T + T$
20
18
16
14
12
10
8
6
4
2
0
0 1 2 3 4 5 6 7 8 9 10
T
(a)
A Sequence
Interventional $a_T + C$
a_T
$a_T + 1$
$a_T + 2$
$a_T + 3$
$a_T + 4$
20
18
16
14
12
10
8
6
4
2
0
0 1 2 3 4 5 6 7 8 9 10
T
(b)



Prior to the Intervention
$a_T + T$
$(t_1, a_{t_1} + t_1)$
$(t_0, a_{t_0} + t_0)$
ΔY
T
After the Intervention
Interventional $a_T + C$
a_T
$a_T + 1$
$a_T + 2$
$a_T + 3$
$a_T + 4$
$T = t_0$
$T = t_1$
ΔY
T

(c) (d)