On the Use of Propensity Scores for Causal Inference with Functional Outcomes and Binary Exposures

Marcelo M. Taddeo and Raydonal Ospina

September 22, 2023

1 Introduction

......

2 General Formulation

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $Y : \Omega \to \mathbb{H}$ be some random functional outcome we are interested in. Here, $(\mathbb{H}, \mathcal{B})$ is a Hilbert (function) space equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{H})$. Possibly the most obvious choice for H is the space of square-integrable functions $L^2[0, 1]$, though other choices are clearly available. We aim to assess the causal effect of an exposure A on the functional outcome Y. For the sake of simplicity, we shall assume A to be binary, say, $A = 0$ (control) or $A = 1$ (treatment). As usual, causal effects are measured through the distribution of the potential outcomes Y^a, $a \in \{0, 1\}$, defined as the outcome functional it would be observed if the exposure were equal to a. Since exposure is determined for each unit, potential outcomes corresponding to unobserved exposures are necessarily counterfactual. This sets up the fundamental problem of causal inference (Holland, 1986).

When considering scalar exposure, the causal effects are generally reported through average causal effects such as the ATE or ATT, **?**. However, if the outcome belongs to some function space, the very notion of causal effects must be reformulated. For example, if $\mathbb{H}$ is a space of functions defined on some interval $\mathcal{I}$, one can use the functional expected causal effect (FATE) proposed by Ecker et al. (2023), given by the stochastic process

$$\tau_{\text{FATE}} = \{\mathrm{E}[Y^1(t) - Y^0(t)] : t \in \mathcal{I}\}.$$

2.1 Identification of the Causal Effect

In order to identify the functional average causal effects, it is enough to consider the following standard assumptions.

Assumption 1 (SUTVA)**.**

$$Y(t) = AY^1(t) + (1 - A)Y^0(t), \quad \forall t \in \mathcal{I}.$$

Assumption 2 (Ignorability)**.**

$$Y^a(t) \perp\!\!\!\perp A|\boldsymbol{X}, \quad \forall t \in \mathcal{I}.$$

....

2.2 Estimation

....

3 Other Stuff

3.1 Banach and Hilbert Space Valued Random Elements

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a given complete probability space and $\mathbb{B}$ a separable Banach space with norm $\|\cdot\|_{\mathbb{B}}$ and equipped with the Borel σ-field $\mathcal{B}(\mathbb{B})$. Since we are primarily concerned with $\mathbb{B}$-valued stochastic processes, we summarize the main concepts we will need.

Definition 1. *A map $X : \Omega \to \mathbb{B}$ is said to be a weakly measurable function if $\psi(X) : \Omega \to \mathbb{R}$ is measurable for every $\psi \in \mathbb{B}'$, where $\mathbb{B}'$ is the topological dual of $\mathbb{B}$. Such functions will be referred to as $\mathbb{B}$-random elements.*

The assumption that X is weakly measurable allows us to consider the following (weak) integral for $\mathbb{B}$-valued functions.

Definition 2. *Any weakly measurable function $X : \Omega \to \mathbb{B}$ satisfying (i) $\psi(X) \in L^1(\Omega, \mathcal{F}, \mathrm{P})$, for all $\psi \in \mathbb{B}'$, and (ii) there is $Y \in \mathbb{B}$ such that, for all $\psi \in \mathbb{B}'$, $\psi(Y) = \int \psi(X) d\,\mathrm{P}$, is said to be weakly integrable. The random element Y is denoted by $\int X d\,\mathrm{P}$ and referred to as the weak integral of X.*

Additionally, if $\mathrm{E}\,\|X\|_{\mathbb{B}} < \infty$, we say that X is integrable. In particular, if X is integrable, then X is also strongly (Bochner) integrable. Actually, if X is integrable, then the notions of weakly and strongly integrable functions are equivalent. This follows from the fact that every integrable X can be approximated by a sequence of simple functions. The reason we use weak integrals instead is that they are more suitable to deal with the second moments of $\mathbb{B}$-random elements. Since P is a probability measure it is natural to identify the weak integral of X with its (weak) expectation, which will be denoted as usual by $\mathrm{E}\,X$. The expectation $\mathrm{E}\,X$ is associated with the operator $\mathrm{E}_X\,\psi = \int \psi(X) d\,\mathrm{P}$, for $\psi \in \mathbb{B}'$, so that $\mathrm{E}\,X$ is such that $\psi(\mathrm{E}\,X) = \mathrm{E}_X\,\psi$, for all $\psi \in \mathbb{B}'$.

Proposition 1. *For any $X \in \mathbb{B}$, the operator $\mathrm{E}_X : \mathbb{B}' \to \mathbb{R}$ is linear and* $\| \mathrm{E}_X \| = \mathrm{E} \|X\|$.

Proof. Linearity follows trivially from the definition of E_X. Now, if $\psi \in \mathbb{B}'$,

$$|\mathrm{E}_X \psi| = |\mathrm{E}\, \psi(X)| \leq \mathrm{E}\, |\psi(X)| \leq \mathrm{E}\, \|\psi\| \|X\| = (\mathrm{E}\, \|\psi\|)\, \|X\|,$$

so that $\| \mathrm{E}_X \| \leq \mathrm{E}\, \|\psi\|$. On the other hand, let $\mathbb{B}_X = \langle X \rangle$, the subspace generated by $X \in \mathbb{B}$, and consider the functional (defined on $\mathbb{B}_X$) $\tilde{\psi}(\alpha X) = \alpha \|X\|$, which clearly is bounded linear with norm $\|\tilde{\psi}\| = 1$. Using the Hahn-Banach Theorem, $\tilde{\psi}$ can be extended to a bounded linear functional ψ on $\mathbb{B}$ such that $\|\psi\| = \|\tilde{\psi}\| = 1$. In particular, $\psi(X) = \tilde{\psi}(X) = \|X\|$. Hence,

$$\| \mathrm{E}_X \| \geq |\mathrm{E}_X \psi| = \left| \int \psi(X) d\, \mathrm{P} \right| = \left| \int \|X\| d\, \mathrm{P} \right| = \mathrm{E}\, \|X\|,$$

so that $\| \mathrm{E}_X \| = \mathrm{E}\, \|X\|$. □

In other words, $\mathrm{E}_X \in \mathbb{B}''$, the topological second dual of $\mathbb{B}$. A similar argument as in the proof of Proposition 1 shows us that $\| \mathrm{E}_X \| = \| \mathrm{E}\, X\|$. For Hilbert spaces $\mathbb{H}$, it follows from Riesz's theorem that there exists an unique $Y \in \mathbb{H}$ such that $\psi(X) = \langle X, Y \rangle$, for every $X \in \mathbb{H}$. Hence, $\mathrm{E}\, X$ is the element in $\mathbb{H}$ such that $\langle \mathrm{E}\, X, Y \rangle = \int \langle X, Y \rangle d\, \mathrm{P}$, for all $Y \in \mathbb{H}$. Moreover, $\mathrm{E}\, X$ is also the representer of E_X, since $\mathrm{E}_X \psi = \langle \mathrm{E}\, X, Y \rangle$, where $Y \in \mathbb{H}$ is the representer of ψ. Finally, notice that $\int [\psi(X - \mathrm{E}\, X)] d\, \mathrm{P} = \int \psi(X) - \psi(\mathrm{E}\, X) = 0$, for all $\psi \in \mathbb{B}'$, so that $\mathrm{E}[X - \mathrm{E}\, X] = 0$. Hence, without loss of generality, we may assume $\mathrm{E}\, X = 0$.

To define an analogous of a covariance measure for Banach space valued random elements, we consider a bilinear version of E_Φ. Given any pair of random elements $X_1 \in \mathbb{B}_1$, $X_2 \in \mathbb{B}_2$, with $\mathrm{E}\, X_1 = 0$, $\mathrm{E}\, X_2 = 0$ and $\psi(X_i) \in L^2(\Omega, \mathcal{F}, \mathrm{P})$, for any $\psi \in \mathbb{B}_i'$, define the covariance functional $\mathrm{C}_{X_1,X_2}(\psi_1, \psi_2) = \int \psi_1(X_1)\psi_2(X_2) d\, \mathrm{P} = \mathrm{Cov}(\psi_1(X_1), \psi_2(X_2))$ on $\mathbb{B}_1' \times \mathbb{B}_2'$. The functional C_{X_1,X_2}

is clearly linear with $|\,\mathrm{C}_{X_1,X_2}(\psi_1,\psi_2)| \leq \mathrm{E}[\|X_1\|\|X_2\|]\|\psi_1\|\|\psi_2\|$ and it is connected to E_X in the following sense: $\mathrm{C}_{X_1,X_2}(\psi_1,\psi_2) = \int \psi_1(\psi_2(X_2)X_1) d\,\mathrm{P} = \mathrm{E}_{\psi_2(X_2)X_1}\,\psi_1 = \mathrm{E}_{\psi_1(X_1)X_2}\,\psi_2$. Conversely, given $\psi_2 \in \mathbb{B}_2'$, $\mathrm{E}_{\psi_1(X_1)X_2}\,\psi_2 = \psi_2(\mathrm{E}[\psi_1(X_1)X_2])$, meaning that $\mathrm{C}_{X_1,X_2}(\psi_1,\psi_2)$ is completely determined by the cross-covariance operators $c_{X_1,X_2} : \psi \in \mathbb{B}_1' \mapsto \mathrm{E}[\psi(X_1)X_2] \in \mathbb{B}_2$. If $X_1 = X_2 = X \in \mathbb{B}$, we write $\mathrm{C}_{X_1,X_2}(\psi_1,\psi_2) = \mathrm{C}_X(\psi_1,\psi_2)$. In this case, the covariance functional is also symmetric positive semidefinite and, by the Schwarz inequality, $|\,\mathrm{C}_X(\psi_1,\psi_2)|^2 \leq \mathrm{V}_X(\psi_1)\,\mathrm{V}_X(\psi_2)$, where $\mathrm{V}_X(\psi) \equiv \mathrm{C}_X(\psi,\psi)$, called the variance functional for X. Both $C_X(\psi,\psi)$ and $V_X(\psi)$ are completely determined by $c_X(\psi) \equiv c_{X,X}(\psi) = \mathrm{E}[\psi(X)X]$ in $\mathbb{B}$. For Hilbert spaces, all these functionals can be clearly written in terms of inner products.

References

Bosq, D. (2000). *Linear Processes in Function Spaces – Theory and Applications*. New York: Springer-Verlag.

Eichler, M. (2012). Causal inference in time series analysis. In Berzuini, C., Dawid, A.P., and Bernardinelli, L. (eds.), *Causality*, 327 – 354. John Wiley and Sons.