A package for a Causal Framework for Hierarchical Outcome Analysis

Francisco Andrade *   Mathieu Even†   Julie Josse‡

Résumé

Quantifying causal effects in the presence of complex and multivariate outcomes remains a major challenge in treatment evaluation. For hierarchical multivariate outcomes, regulatory agencies such as the U.S. Food and Drug Administration recommend approaches based on the Win Ratio and Generalized Pairwise Comparisons. However, commonly used estimators typically target a population-level estimand—the probability that a randomly sampled treated patient fares better than a randomly sampled control patient. This estimand may lead to treatment recommendations that differ from those based on an ideal causal estimand : the probability that a given individual would fare better under treatment than under control. In heterogeneous populations, this discrepancy highlights both the non-identifiability of the individual-level estimand and the importance of clearly specifying the underlying causal framework.

In this context, we introduce an R package implementing the methodology proposed in Even and Josse [2025]. The package provides tools to estimate a causal effect measure designed to better approximate the ideal individual-level estimand. It implements three complementary estimators : (i) a nearest-neighbor matching win ratio estimator, (ii) a distributional regression estimator, and (iii) a semiparametric efficient estimator. By providing a unified and accessible implementation of these methods, the package facilitates practical application of causal inference techniques for hierarchical outcome analysis in clinical and biomedical research.

Mots-clefs (3 à 5) : Statistique – Biostatistique – Package

Développement

Setting Consider n i.i.d. samples (X_i, T_i, Y_i), where X_i denotes individual covariates, $T_i \in \{0,1\}$ is the treatment indicator, and $Y_i \subseteq \mathcal{Y}$ represents the (possibly multivariate) outcome. Given a win function

$$w : \mathcal{Y} \times \mathcal{Y} \to [0,1],$$

our package provides implementations of the following estimators :

$$\hat{\tau}_{\mathrm{NN}} := \frac{1}{n} \sum_i T_i w(Y_i | Y_{\sigma_1^\star(i)}) + (1 - T_i) w(Y_{\sigma_0^\star(i)} | Y_i), \tag{1}$$

where, for $t \in \{0,1\}$, $\sigma_t^\star$ assigns the index of the closest element in the other group, *i.e.,* group $1 - t$;

$$\hat{\tau}_{\mathrm{reg}} := \frac{1}{n} \sum_i T_i \hat{q}_1(X_i, Y_i) + (1 - T_i) \hat{q}_0(X_i, Y_i), \tag{2}$$

*INRIA, `francisco.de-lima-andrade@inria.fr`
†INRIA
‡INRIA

where $\hat{q}_0$ and $\hat{q}_1$ are estimators of the nuisance parameters $q_0(x,y) = \mathbb{E}[w(y|Y_i)|T_i = 0, X_i = x]$ and $q_1(x,y) = \mathbb{E}[w(Y_i|y)|T_i = 1, X_i = x]$;

$$\begin{aligned}\hat{\tau}_{\mathrm{reg}} :=& \frac{1}{n}\sum_i T_i w(Y_i|Y_{\sigma_1^\star(i)}) + (1-T_i)w(Y_{\sigma_0^\star(i)}|Y_i) \\ &+ \frac{T_i}{\hat{\pi}(X_i)}\Big(\hat{q}_1(X_i,Y_i) - w(Y_i|Y_{\sigma_1^\star(i)})\Big) + \frac{1-T_i}{1-\hat{\pi}(X_i)}\Big(\hat{q}_0(X_i,Y_i) - w(Y_{\sigma_0^\star(i)}|Y_i)\Big),\end{aligned} \tag{3}$$

where $\hat{q}_0, \hat{q}_1$ are as in (2), $\sigma_t^\star$ is as in () and $\hat{\pi}$ is an estimator of the propensity score.

Causal Measure These estimators target the causal measure defined in Even and Josse [2025].

Features The main features of our implementation are :

- Mixed covariates can be handled by transforming them using `FAMD` from the `FactoMineR` package.
- The implementation of (2) accommodates missing values in the covariates by leveraging the `drf` package.
- We implement the confidence intervals for our estimators.
- Both propensity scores and the nuisance parameters are implemented using cross-fitting.

GitHub: https://github.com/FranciscoAndrade90/causalWins

Références

Mathieu Even and Julie Josse. Rethinking the win ratio : A causal framework for hierarchical outcome analysis. *arXiv preprint arXiv :2501.16933*, 2025.