

Implementing bi-level matching optimization with propensity score

For PHS 7045

George G. Vega Yon

Introduction

Matching algorithms for causal inference are a fundamental tool in biomedical sciences. Nonetheless problems about optimal matching strategies, balance, and related issues are still open. In this report, I will introduce an implementation of the bi-level matching optimization algorithm published by XYZ and show how it can be used to improve the balance of covariates in a dataset, particularly, looking into the propensity score matching.

Description of the solution

To address the problem, we will be implementing the algorithm using C++ via Rcpp in R. The algorithm will be implemented in the following steps:

1. Compute propensity scores using a logistic regression in R.
2. Use the propensity scores to do a 1:2 matching for generating the counterfactuals.
3. Return the results of the matching a list with: (a) a two column matrix showing the idx of the matched pairs, and (b) the matched estimator.

Results

```
library(Rcpp)

# Simulating some data
n <- 1000
set.seed(3112)
```

```
group <- sample.int(2, n, replace = TRUE) - 1
X <- runif(n) # Simulating propensity score
Y <- rnorm(n) # simulating outcome
```

```
sourceCpp("match.cpp", cacheDir = "cache")
ans <- match_it(
  X = X,
  group = group,
  Y = Y
)
```

The following figure shows how the match works for the first ten observations. Red and black points should match!

```
# Looking into the first 10
plot(X[1:10])
points(X[ans$matches[1:10, 1] + 1], col = "red", pch=20)
points(X[ans$matches[1:10, 2] + 1], col = "red", pch=15)
points(X[1:10], col = "black")
```

