ATEHonest: Honest CIs for Average Treatment Effects

Michal Kolesár

June 28, 2020

The package `ATEHonest` implements honest confidence intervals and estimators for estimating average treatment effects under unconfoundedness from Armstrong and Kolesár [2018]. Here we illustrate the use of the package using NSW data from Dehejia and Wahba [1999].

The data is shipped with the package, as two data frames, `NSW` (where the treated units are from the experimental sample and control units are from PSID), and `NSWexper`, where both treated and control units are from the experimental sample. We'll use the experimental sample here.

First we extract the design matrix, and the treatment and outcome vectors:

```
library("ATEHonest")
X <- as.matrix(NSWexper[, 2:10])
d <- NSWexper$treated
y <- NSWexper$re78
```

Next, we compute matrix of distances between treated and control units, using the same weight matrix to compute distances as in Armstrong and Kolesár [2018]:

```
Ahalf <- diag(c(0.15, 0.6, 2.5, 2.5, 2.5, 0.5, 0.5, 0.1,
    0.1))
D0 <- distMat(X, Ahalf, method = "manhattan", d)
```

We now compute the first few steps of the solution path. 120 steps will be sufficient for computing the optimal estimator.

```
op <- ATTOptPath(y, d, D0, maxsteps = 120)
```

Next, construct a distance matrix used by the nearest neighbor variance estimator to estimate the conditional variance of the outcome. We use Mahalanobis distance:

```
DM <- distMat(X, chol(solve(cov(X))), method = "euclidean")
```

Given the solution path, we now compute the root mean squared error optimal estimator, as well as estimator that's optimal for constructing two-sided CIs:

```
ATTOptEstimate(op, C = 1, DM = DM, opt.criterion = "RMSE")
#>
#>
#> Estimate   Max. bias   SE          CI                        delta
#> ---------  ----------  ----------  -----------------------  ---------
#> 1.58946    1.185779    0.7550322   (-0.8382422, 4.017163)    1.222475
```

```r
ATTOptEstimate(op, C = 1, DM = DM, opt.criterion = "FLCI")
#>
#>
#> Estimate   Max. bias   SE          CI                          delta
#> ---------  ----------  ---------  -----------------------  ---------
#> 1.62276    1.234971    0.713424   (-0.7856902,  4.03121)    3.289756
```

For computing efficiency of one- and two-sided CIs at smooth functions (see Appendix A in Armstrong and Kolesár [2018]), the solution path is not long enough:

```r
ATTEffBounds(op, DM = DM, C = 1)
#> Warning in ATTEffBounds(op, DM = DM, C = 1): Path too short to compute one-sided
#> efficiency
#> Warning in ATTEffBounds(op, DM = DM, C = 1): Path too short to compute two-sided
#> efficiency
#> $onesided
#> [1] NaN
#>
#> $twosided
#> [1] NaN
```

We therefore make it longer, by passing the output `op` as an argument to `ATTOptPath`

```r
op <- ATTOptPath(path = op, maxsteps = 290)
ATTEffBounds(op, DM = DM, C = 1)
#> $onesided
#> [1] 0.9917693
#>
#> $twosided
#> [1] 0.9750232
```

For comparison, we also consider matching estimators. First, a matching estimator with a single match:

```r
ATTMatchEstimate(ATTMatchPath(y, d, D0, M = 1, tol = 1e-12),
    C = 1, DM = DM)
#>
#>
#> Estimate   Max. bias   SE          CI                          M
#> ---------  ----------  ----------  -----------------------  ---
#> 1.972065   1.169562    0.7758019   (-0.4735895,  4.41772)     1
```

Next, we optimize the number of matches. For that we first compute the matching estimator for a vector of matches `M`, and then optimize the number of matches using `ATTMatchEstmate`:

```r
mp <- ATTMatchPath(y, d, D0, M = 1:10, tol = 1e-12)
ATTMatchEstimate(mp, C = 1, DM = DM, opt.criterion = "FLCI")
#>
#>
#> Estimate   Max. bias   SE          CI                          M
```

```
#> ---------  ----------  ----------  ----------------------  ---
#> 1.972065   1.169562    0.7758019   (-0.4735895,  4.41772)    1
ATTMatchEstimate(mp, C = 1, DM = DM, opt.criterion = "RMSE")
#>
#>
#> Estimate   Max. bias   SE          CI                        M
#> ---------  ----------  ----------  ----------------------  ---
#> 1.972065   1.169562    0.7758019   (-0.4735895,  4.41772)    1
```

We can see that a single match is in fact optimal for both estimation and construction of two-sided CIs.

References

Tim Armstrong and Michal Kolesár. Finite-sample optimal estimation and inference on average treatment effects under unconfoundedness. arXiv: 1712.04594, December 2018. URL https://arxiv.org/abs/1712.04594.

Rajeev H. Dehejia and Sadek Wahba. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448):1053–1062, December 1999. doi: 10.1080/01621459.1999.10473858.