

geoMatch

Description

geoMatch improves models using spatial data for the purposes of causal inference by selecting matched subsets of the original treated and control groups. It provides an extension of the R package MatchIt (see Ho, Imai, King, and Stuart (2004)), enabling the use of spatial data frames as well as providing an adjustment factor to mitigate potential spillover between treated and control units (i.e., in cases where stable unit treatment value assumptions (SUTVA) may not be met due to spatial spillovers). geoMatch maintains the full functionality of MatchIt, including all matching strategies and caliper functions. Full documentation on MatchIt is available online at <http://gking.harvard.edu/matchit>, and help for specific commands is available through `help.matchit`.

Usage

```
geoMatch(..., outcome.variable, outcome.suffix = "_adjusted",
  optim.iterations = 10000)
```

Arguments

- `...` The first parameters provided to geoMatch should be a traditional MatchIt specification. Full documentation on MatchIt is available online at <http://gking.harvard.edu/matchit>. Dataframe must be a spatial points dataframe.
- `outcome.variable` The name of the outcome variable that will be modeled to establish causal effect. This must be an existing attribute in the spatial dataframe passed to geoMatch.
- `outcome.suffix` Suffix for the returned column name with spillover-adjusted outcome data.
- `optim.iterations` the number of iterations to perform within the spatial decay optimization procedure.

Details

geoMatch overcomes two challenges. First, it seeks to first remove spillover from control outcome measures, for example a case where a clinic may improve health outcomes in both the geographic neighborhood it is located in, as well as nearby neighborhoods. Failing to adjust for this spillover can result in erroneous estimates of impact. Second, it allows for the use of spatial points data frames in conjunction with the MatchIt matching framework. geoMatch returns an adjusted version of the outcome variable for each control case as specified by the user, defined as Y^* . Y^* can be interpreted as the estimated outcome if the spatial spillovers from treated cases are netted out. This adjusted Y^* can be used in the second stage model for more accurate estimations of treatment effects. Y^* is calculated through a multiple step process. First, a distance matrix (D_{ct}) is constructed which provides the Euclidean distances between each treated and control case:

```
[Dc1t1 Dc1t2 Dc1t3 . Dc1tn]
[ . . . . . ]
[ . . . . . ]
[Dcnt1 Dcnt2 Dcnt3 . Dcntn]
```

Second, two vector (Y_t and Y_c) are constructed which contain the known, observed outcomes in both Y_t and Y_c . Third, a spherical distance decay function is fit for each T simultaneously, in which the parameter U_t is solved for across all units T in the spatial function `sf(distance, U)`:

$$(1 - [3/2] * (D_{ct} / U_t) - [1/2] * (D_{ct}/U_t)^3)$$

where we seek to optimize the absolute difference between the observed outcome at each control location (Y_c) and the estimated outcome (Y_{c_e}) as a product of neighboring units Y_t :

```
Yc_e = sf(Dct, Ut) * Yt
minimize(abs(Yc_e-Yc))
```

The vector of distances U_t is used to estimate the adjusted Y_c^* , which - for each control case - provides an estimate of the outcome if spillover is removed:

$$Y_c^* = Y_c - (sf[Dct, U_t] * Y_t).$$

geoMatch then proceeds as usual with MatchIt, returning a matched set of spatial locations alongside Y_c^* .

Value

This function will return a MatchIt object, with a spatial data frame accesible in \$spdf.

Authors

Dan Miller Runfola dan@danrunfola.com; Ariel BenYishay abenyishay@wm.edu; Seth Goodman sgoodman@aiddata.org

Citation

If you find this package useful, please cite:
Runfola, D., BenYishay, A., Goodman, S., 2016. geoMatch: An R Package for Spatial Propensity Score Matching. R package. <http://geo.aiddata.org>.

Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political Analysis 15(3): 199-236.
<http://gking.harvard.edu/files/abs/matchp-abs.shtml>

Source

<http://geo.aiddata.org/>

Examples

```
library(geoMatch)

###
### An Example Script for Obtaining Matched Data when you have
### Spatial information
###
data(lalonde)

##Traditional, non-spatial MatchIt
library(MatchIt)
match.out1 <- matchit(treat ~ age + educ + black + hispan +
                     nodegree + married + re74 + re75,
                     method = "nearest", data = lalonde)

##Example model performed after matching, including both Control and Treatment groups
lm.out1 <- lm(re78 ~ treat + age + educ + black + hispan +
             nodegree + married + re74 + re75 + distance,
             data = match.data(match.out1))
summary(lm.out1)

##Simulate Latitude and Longitude information for each point,
##with enforced spatial correlation.
library(sp)
set.seed(500)
```

```

coords = cbind(runif(nrow(lalonde),37.1708,37.3708),
               runif(nrow(lalonde),76.6069,76.8069))

##Create a spatial points data frame
spatial_lalonde <- SpatialPointsDataFrame(coords, lalonde)

##Matching and adjusting for spillover effects
##See ?geoMatch for more parameters specific to spatial data.
##See ?MatchIt for more options for matching methods.
match.out2 <- geoMatch(treat ~ age + educ + black + hispan + nodegree +
                      married + re74 + re75,
                      method = "nearest",
                      caliper=0.25,
                      data = spatial_lalonde,
                      outcome.variable="re78",
                      outcome.suffix="_adjusted",
                      optim.iterations = 100)

##Example maps
spplot(match.out2$spdf, z="matched", col.regions=c("red","green"),
        main="Map of Matched Pairs")
spplot(match.out2$spdf, z="distance", main="Propensity Scores")
spplot(match.out2$spdf, z="est_spillovers", main="Estimated Spillovers")
spplot(match.out2$spdf, z="re78_adjusted", main="Adjusted Outcome")

#Percent of outcomes attributable to spillovers
match.out2$spdf@data["spill_percent"] <- 100 *
  (match.out2$spdf@data["est_spillovers"] / match.out2$spdf@data["re78"])

spplot(match.out2$spdf[!is.infinite(match.out2$spdf@data$spill_percent) &
                      match.out2$spdf@data$treat == 0,],
        z="spill_percent",
        main="% Outcome Attributable to Spillover",
        pretty=TRUE,
        cuts=5)

##Example model performed after spatial spillover adjustment, using matched data
lm.out2 <- lm(re78_adjusted ~ treat + age + educ + black + hispan + nodegree +
             married + re74 + re75 + distance,
             data = match.data(match.out2))
summary(lm.out2)

##Example model with spatial lag after spatial spillover adjustment
##For more information on practical specifications for these models, see
##Corrado, Luisa, and Bernard Fingleton. Where is the economics in spatial
##econometrics?. Journal of Regional Science 52.2 (2012): 210-239.
library(spdep)
matched.spatial <- match.out2$spdf[match.out2$spdf@data$matched == 1,]
coords <- coordinates(matched.spatial)
k1 <- knn2nb(knearneigh(coords, k=15))
all.linked <- max(unlist(nbdists(k1, coords)))
neighbor.list <- dnearneigh(coords, 0, all.linked)
sl.out3 <- lagsarlm(re78 ~ treat + age + educ + black + hispan +
                  nodegree + married + re74 + re75 + distance,
                  data=matched.spatial,
                  nb2listw(neighbor.list, style="W"), method="MC", quiet=TRUE,
                  tol.solve=1.0e-16)
summary(sl.out3)

```

[Package *geoMatch* version 0.1.1]