IPW models

TingFang, URI Avenir Team

```
library(lme4)
library(plyr)
library(dplyr)
library(igraph)
library(numDeriv)
library(gtools)
rm(list = ls())

source("https://github.com/uri-ncipher/Nearest-Neighbor-estimators/blob/main/functions.R?raw=TRUE")
```

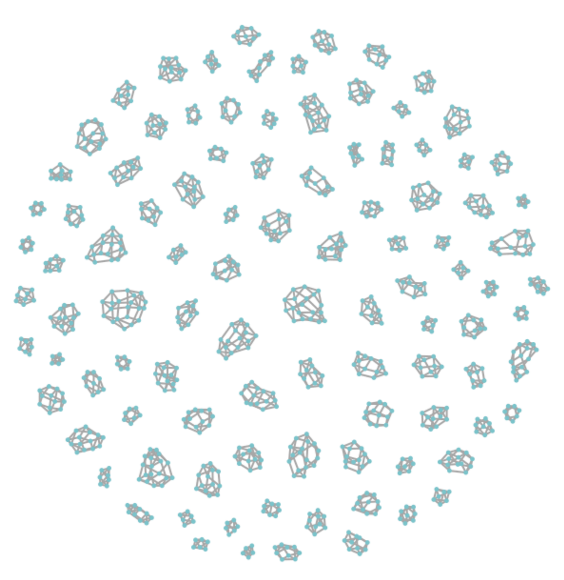
Data Preparation

Read in the synthetic nodes and edges data sets for creating the simulated network.

```
nodes=read.csv("https://github.com/uri-ncipher/Nearest-Neighbor-estimators/blob/main/nodes.csv?raw=TRUE")
edges=read.csv("https://github.com/uri-ncipher/Nearest-Neighbor-estimators/blob/main/edges.csv?raw=TRUE")
net0=graph_from_data_frame(d=edges, vertices = nodes, directed=F)
```

Network visualization

```
# plot the network generated above
plot(net0, vertex.size=1, vertex.label=NA, vertex.color="cadetblue3", vertex.frame.color="cadetblue3")
```



Make the data for modeling

```
# Save the number of individuals (nodes) and number of components
n=length(V(net0))      # number of individuals
m=components(net0)$no  # number of components

# generate a data set with three columns, the first column is participant id (node id),
# the second column is number of nearest neighbors for each participant (is also node degree),
# the third column indicates the component for each participant.
data=data.frame(id=1:n, na=unlist(lapply(1:n, num_neighbors, net=net0)),
                component=components(net0)$membership)

# assign treatment, outcome, and baseline covariates to the dataset 'data'.
data$treatment=nodes$treatment
data$outcome=nodes$outcome
data$var1=nodes$var1
data$var2=nodes$var2
data$na_a= unlist(lapply(1:n, trt_neighbors, net=net0))
data$notna_a=data$na-data$na_a

base_covariate=c("var1", "var2")

# averaging baseline covariates (average of nearest neighbors' baseline covariate values)
data$avg_var1=unlist(lapply(1:n, avg_neighbors, net=net0, variable="var1"))
data$avg_var2=unlist(lapply(1:n, avg_neighbors, net=net0, variable="var2"))

avg_covariate=c("avg_var1", "avg_var2")
```

Modeling

In this section, we will apply both IPW1 and IPW2 estimators for the average potential outcomes and corresponding causal effects. An individual's exposure/treatment and allocation strategy α jointly determine the average potential outcome (allocation strategy represents the probability that individuals in nearest neighbors set receiving the exposure/treatment in the counterfactual scenario. so it is a numeric value between 0 and 1). For each estimator, the R programs will output the **point estimation and the estimated variance of average potential outcomes** $\boldsymbol{\widehat{Y}(1,\alpha), \widehat{Y}(0,\alpha), \widehat{Y}(\alpha)}$ under three different allocation strategies α (i.e., 0.25, 0.50, and 0.75), respectively. Also, the R program will output the **point estimation and estimated variance of four causal effects: Direct** (DE), **Indirect** (IE), **Total** (TE), and **Overall effect** (OE).

Equations for calculating the four causal effects are:

- $\widehat{DE} = \widehat{Y}(1,\alpha) - \widehat{Y}(0,\alpha)$
- $\widehat{IE} = \widehat{Y}(0,\alpha_0) - \widehat{Y}(0,\alpha_1)$
- $\widehat{TE} = \widehat{Y}(1,\alpha_0) - \widehat{Y}(0,\alpha_1)$
- $\widehat{OE} = \widehat{Y}(\alpha_0) - \widehat{Y}(\alpha_1)$,

where α_0, α_1 both represents allocation strategies, and $\alpha_0 \neq \alpha_1$ (Note: The associated paper compared the estimated average potential outcomes with α_1 to α_0 for the indirect, total, and overall effects. The code here compared the estimated average potential outcomes with α_0 to α_1 for the indirect, total, and overall effects). Users can set any values between 0 and 1 for α_0 and α_1. If given α is a vector, then the causal contrasts will produce pairwise comparisons in the sequential order of the values in α. For example, if $\alpha = c(0.75, 0.25, 0.50)$, the contrasts for spillover effects will compare

- i. $\alpha_0 = 0.75$ versus $\alpha_1 = 0.25$;
- ii. $\alpha_0 = 0.75$ versus $\alpha_1 = 0.50$;
- iii. $\alpha_0 = 0.25$ versus $\alpha_1 = 0.50$.

IPW1

Using IPW1 to estimate the average potential outcomes and causal effects under allocation strategies α.

```
alpha=c(0.25, 0.50, 0.75) # set allocation strategies
IPW_1_model(data, base_covariate, alpha)  # call on IPW1 function
```

```
## [[1]]
##             a=1         a=0       margin alpha           type
## 1 0.1852175678 0.188372738 0.1875839458  0.25 point estimate
## 2 0.1451336794 0.276759930 0.2109468049  0.50 point estimate
## 3 0.1352103897 0.261785879 0.1668542621  0.75 point estimate
## 4 0.0014465618 0.000919702 0.0005100593  0.25       variance
## 5 0.0003415270 0.001480020 0.0004108478  0.50       variance
## 6 0.0004228426 0.001566447 0.0003414386  0.75       variance
##
## [[2]]
##       estimation alpha0 alpha1     type
## 1  -0.0031551706   0.25   0.25   Direct
## 2  -0.1316262511   0.50   0.50   Direct
## 3  -0.1265754895   0.75   0.75   Direct
## 4   0.0028872409   0.25   0.25   Var DE
## 5   0.0019997039   0.50   0.50   Var DE
## 6   0.0019589605   0.75   0.75   Var DE
## 7  -0.0883871920   0.25   0.50 Indirect
## 8  -0.0734131408   0.25   0.75 Indirect
## 9   0.0149740512   0.50   0.75 Indirect
## 10  0.0004993204   0.25   0.50   Var IE
## 11  0.0018240734   0.25   0.75   Var IE
## 12  0.0012634087   0.50   0.75   Var IE
## 13 -0.0915423626   0.25   0.50    Total
## 14 -0.0765683114   0.25   0.75    Total
## 15 -0.1166521998   0.50   0.75    Total
## 16  0.0033180932   0.25   0.50   Var TE
## 17  0.0024878018   0.25   0.75   Var TE
## 18  0.0017330800   0.50   0.75   Var TE
## 19 -0.0233628591   0.25   0.50  Overall
## 20  0.0207296837   0.25   0.75  Overall
## 21  0.0440925428   0.50   0.75  Overall
## 22  0.0001937051   0.25   0.50   Var OE
## 23  0.0006692608   0.25   0.75   Var OE
## 24  0.0003768236   0.50   0.75   Var OE
```

In the output above,

- The **section [[1]]** displays the point estimates (type = "point estimate" in the output) and estimated variances (type = "variance" in the output) for the average potential outcomes under three different allocation strategies (i.e., $\alpha = 0.25, 0.50, 0.75$) and each individual exposure ($a = 0, a = 1$, and margin). "Margin" is the average potential outcomes for a particular allocation strategy, regardless of individual exposure status.
- The **section [[2]]** displays the point estimates (type = "Direct", "Indirect", "Total", "Overall" in the output) and estimated variances (type = "Var DE", "Var IE", "Var TE", "Var OE" in the output) of four causal effects: direct, indirect, total and overall effects, corresponding to particular allocation strategies (α_0 and α_1). The estimated values are shown in "estimation" column in the output above.

IPW2

Using IPW2 to estimate the average potential outcomes and causal effects under allocation strategies α.

```
# note: the same allocation strategies applied here. i.e., alpha=c(0.25, 0.5, 0.75)

# fit logistic regression models for nearest neighbors' exposures (M1) and individual's exposure (M2), respective
ly.
formula_1=paste("cbind(na_a, notna_a)", "~", "treatment", "+",
                paste(base_covariate, collapse = "+"), "+",
                paste(avg_covariate, collapse = "+"))    # formula for fitting M1
formula_2=paste("treatment", "~", paste(base_covariate, collapse = "+")) # formula for fitting M2
M1=glm(formula_1, family = binomial(link = "logit"), data=data)
M2=glm(formula_2, family = binomial(link = "logit"), data=data)

IPW_2_model(data, M1, M2, alpha) # call on IPW2 function
```

```
## [[1]]
##             a=1          a=0       margin alpha           type
## 1 0.1796249651 0.1454643863 0.1540045310  0.25 point estimate
## 2 0.1597258863 0.2266322694 0.1931790778  0.50 point estimate
## 3 0.1470394052 0.2526265019 0.1734361794  0.75 point estimate
## 4 0.0011119881 0.0003719574 0.0002913227  0.25       variance
## 5 0.0003355471 0.0004879431 0.0002299846  0.50       variance
## 6 0.0004107754 0.0016837676 0.0003741267  0.75       variance
##
## [[2]]
##       estimation alpha0 alpha1     type
## 1   0.0341605788   0.25   0.25   Direct
## 2  -0.0669063830   0.50   0.50   Direct
## 3  -0.1055870966   0.75   0.75   Direct
## 4   0.0014167592   0.25   0.25   Var DE
## 5   0.0007270418   0.50   0.50   Var DE
## 6   0.0018927829   0.75   0.75   Var DE
## 7  -0.0811678831   0.25   0.50 Indirect
## 8  -0.1071621155   0.25   0.75 Indirect
## 9  -0.0259942325   0.50   0.75 Indirect
## 10  0.0002295839   0.25   0.50   Var IE
## 11  0.0017476652   0.25   0.75   Var IE
## 12  0.0010428090   0.50   0.75   Var IE
## 13 -0.0470073043   0.25   0.50    Total
## 14 -0.0730015368   0.25   0.75    Total
## 15 -0.0929006155   0.50   0.75    Total
## 16  0.0014390786   0.25   0.50   Var TE
## 17  0.0020005766   0.25   0.75   Var TE
## 18  0.0015989317   0.50   0.75   Var TE
## 19 -0.0391745468   0.25   0.50  Overall
## 20 -0.0194316484   0.25   0.75  Overall
## 21  0.0197428985   0.50   0.75  Overall
## 22  0.0001526009   0.25   0.50   Var OE
## 23  0.0005000172   0.25   0.75   Var OE
## 24  0.0002177706   0.50   0.75   Var OE
```

In the output above,

- The **section [[1]]** displays the point estimates (type = "point estimate" in the output) and estimated variances (type = "variance" in the output) for the average potential outcomes under three different allocation strategies (i.e., $\alpha = 0.25, 0.50, 0.75$) and each individual exposure ($a = 0, a = 1$, and margin). "Margin" is the average potential outcomes for a particular allocation strategy, regardless of individual exposure status.
- The **section [[2]]** displays the point estimates (type = "Direct", "Indirect", "Total", "Overall" in the output) and estimated variances (type = "Var DE", "Var IE", "Var TE", "Var OE" in the output) of four causal effects: direct, indirect, total and overall effects, corresponding to particular allocation strategies (α_0 and α_1). The estimated values are shown in "estimation" column in the output above.