Package 'aggregation'

October 28, 2020

Type Package

Title R package for aggregation of device numbers into population number

Version 0.1.0

Author Bogdan Oancea <bogdan.oancea@gmail.com>, David Salgado <david.salgado.fernandez@ine.es> Sandra Barragan <sandra.barragan.andres@ine.es>

Maintainer Bogdan Oancea <bogdan.oancea@gmail.com>

Description R package for aggregation of the number of devices detected by mobile network into the number of individuals.

License GPL3, EUPL

Imports data.table,
deduplication,
Matrix,
doParallel,
parallel,
extraDistr

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Suggests knitr,
rmarkdown

VignetteBuilder knitr

Collate 'aggregation.R'
'buildCluster.R'
'buildPostLocJointProbs.R'
'rNnet_Event.R'
'doAggr.R'
'example.R'
'nIndividuals.R'
'nIndividuals3.R'
'rNnetEvent.R'
'rNnetJoint_Event.R'
'rNnetEventOD.R'

R topics documented:

`aggregation` *aggregation: A package for computing the number of individuals detected by the network.*

Description

This package contains functions to compute the number of individuals detected by the network in each geogrphical region as well as the number of individuals detected by the network moving from one region to another at successive time instants. For an example on how to use this package please read example.

`example` *Example of using the aggregation package*

Description

This is just an example on how to use this package to generate random values from a Poisson multinomial distribution in order to obtain a point estimate of the number of individuals detected by mobile network in a region.

Usage

```
example()
```

Details

This is a script that shows how to use the functions of this package to compute a point estimate of the number individuals detected by the network in each region and the number of individuals moving from one region to another. From the set of random values one can obtain any point estimate: mean. mode, median.

References

https://github.com/MobilePhoneESSnetBigData

Examples

```
# set the folder where the necessary input files are stored and the prefix of
# the input file names.
path      <- 'extdata'

prefix='postLocDevice'

# gets the series of time instants from the simulation.xml file.
simParams <-deduplication::readSimulationParams(system.file(path,
'simulation.xml', package = 'aggregation'))
time_from <- simParams$start_time
time_to <- simParams$end_time
time_incr <- simParams$time_increment
times<-seq(from=time_from, to=time_to-time_incr, by = time_incr)

# set the grid file name, i.e. the file the parameters of the grid
grFile <- system.file(path, 'grid.csv', package = 'aggregation')

# set the duplicity probabilities file name, i.e. the file with duplicity
# probability for each device
dpFile<-system.file(path, 'duplicity.csv', package = 'aggregation')

# set the regions file name, i.e. the file defining the regions for which we
# need the estimation of the number of individuals detected by network.
rgFile<-system.file(path, 'regions.csv', package = 'aggregation')

# generate n random values
n <- 1e3
nNet <- rNnetEvent(n, grFile, dpFile, rgFile, system.file(path,
package = 'aggregation'), prefix, times = times)

# print the mean number of detected individuals for each region, for each
# time instant
regions <- as.numeric(unique(nNet$region))
times <- unique(nNet$time)

for(r in regions) {
    print(paste0("region: ", r))
    for(t in times) {
        print(paste0("time instant: ", t, " number of individuals: " ,
        round(mean(nNet[region == r][time ==t]$N))))
    }
}


# For the origin-destination matrix we proceed in a similar way
prefixJ <- 'postLocJointProbDevice'
nnetOD <- rNnetEventOD(n, dpFile, rgFile, system.file(path,
package = 'aggregation'), prefixJ)

# The origin-destination matrix can be computed now very simple
# First we choose two consecutive time instants
t1 <- 0
t2 <- 10
# The we extract the regions:
```

```
regions_from <- sort(as.numeric(unique(nnetOD$region_from)))
regions_to <- sort(as.numeric(unique(nnetOD$region_to)))

# Now we compute the origin-destination matrix:
ODmat <- matrix(nrow = length(regions_from), ncol = length(regions_to))
for(r1 in regions_from) {
   for(r2 in regions_to) {
       ODmat[r1,r2] <-
       round(mean(nnetOD[time_from==t1][time_to==t2][region_from==r1][region_to==r2]$Nnet))
   }
}
ODmat
```

`rNnetEvent` *Generates random values according to a Poisson multinomial probability distribution.*

Description

Generates random values according to a Poisson multinomial probability distribution. A point estimation derived from this distribution (mean, mode) represents an estimation of the number of individuals detected by the network in a region. Regions are composed as a number of adjacent tiles. This is the only one function of this package available to users to compute an estimation of the number of detected individuals. For a theoretical background an interested reader can find more details in the methodological framework available here: https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/f/fb/WPI_Deliverable_I3_A_proposed_production_framework_with_mobile_network_data_2020_05_31_draft.pdf

Usage

```
rNnetEvent(
  n,
  gridFileName,
  dupFileName,
  regsFileName,
  postLocPath,
  prefix,
  times,
  seed = 123
)
```

Arguments

`n`	The number of random values to be generated.
`gridFileName`	The name of the .csv file with the grid parameters.
`dupFileName`	The name of the .csv file with the duplicity probability for each device. This is an output of the `deduplication` package.

regsFileName	The name of the .csv file defining the regions. It has two columns: tile,region. The first column contains the IDs of each tile in the grid while the second contains the number of a region. This file is defined by the user and it can be created with any text editor.
postLocPath	The path where the files with the posterior location probabilities for each device can be found. A file with the location probabilities should have the name prefix_ID.csv where ID is replaced with the device ID and prefix is given as a parameter to this function.
prefix	A prefix that is used to compose the file name with posterior location probabilities.
times	A vector with the time instants when the events were registered.
seed	The value of the random seed to be used by the random number generator.

Value

A data.table object with the following columns: time,region,N,iter. The last column contains the index of the random value (given in column N) generated for each time instant and region.

---

rNnetEventOD *Generates random value according to a Poisson multinomial distribution needed to estimate the origin destination matrices.*

Description

Generates random value according to a Poisson multinomial distribution needed to estimate the origin destination matrices. This is a high level function, the only one to be called by users to estimate the number of individuals going from one region to another. For a theoretical background an interested reader can find more details in the methodological framework available here: https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/f/fb/WPI_Deliverable_I3_A_proposed_production_framework_with_mobile_network_data_2020_05_31_draft.pdf The actual omputations are performed using a parallelization (transparent to the users) which uses the whole number of (logical) cores.

Usage

```
rNnetEventOD(
  n,
  dupFileName,
  regsFileName,
  postLocJointPath,
  prefix,
  seed = 123
)
```

Arguments

n	The number of random values to be generated.
dupFileName	The name of the .csv file with the duplicity probability for each device. This is an output of the deduplication package.

`regsFileName` The name of the .csv file defining the regions. It has two columns: `tile,region`. The first column contains the IDs of each tile in the grid while the second contains the number of a region. This file is defined by the user and it can be created with any text editor.

`postLocJointPath` The path where the files with the posterior location probabilities for each device can be found. A file with the location probabilities should have the name `prefix_ID.csv` where `ID` is replaced with the device ID and `prefix` is given as a parameter to this function.

`prefix` A prefix that is used to compose the file name with posterior location probabilities.

Value

A data table object with the following columns: `time_from,time_to,region_from,region_to,Nnet,iter`. The number of detected individuals moving from a region to another between two succesive time instants is given in column `Nnet` while the last column gives the index of the random value generated for this number.

Index