Package 'evidenceFactors'

February 20, 2020

Type Package

Title Reporting Tools for Sensitivity Analysis of Evidence Factors in
Observational Studies

Version 1.8

Date 2020-02-14

Author Bikram Karmakar

Maintainer Bikram Karmakar <bkarmakar@ufl.edu>

Description Provides tools for integrated sensitivity analysis of evidence factors in observational
studies. When an observational study allows for multiple independent or nearly
independent inferences which, if vulnerable, are vulnerable to different biases, we have
multiple evidence factors. This package provides methods that respect type I error rate control.
Examples are provided of integrated evidence factors analysis in a longitudinal study with
continuous outcome and in a case-control study.
Karmakar, B., French, B., and Small, D. S. (2019)<DOI:10.1093/biomet/asz003>.

License MIT + file LICENSE

Imports sensitivitymv

NeedsCompilation no

Repository CRAN

Date/Publication 2020-02-20 15:00:02 UTC

R topics documented:

---

evidenceFactors-package

Reporting Tools for Sensitivity Analysis of Evidence Factors in Observational Studies

---

Description

Provides tools for integrated sensitivity analysis of evidence factors in observational studies. When an observational study allows for multiple independent or nearly independent inferences which, if vulnerable, are vulnerable to different biases, we have multiple evidence factors. This package provides methods that respect type I error rate control. Examples are provided of integrated evidence factors analysis in a longitudinal study with continuous outcome and in a case-control study. Karmakar, B., French, B., and Small, D. S. (2019)<DOI:10.1093/biomet/asz003>.

Details

Index of help topics:

```
Gamlist                 Gamma list for illustration of the use of the
                        package.
Plist                   p-value list for illustration of the use of the
                        package.
evidenceFactors-package
                        Reporting Tools for Sensitivity Analysis of
                        Evidence Factors in Observational Studies
mtm                     DNA damage from exposure to chromium
plotRejDecbyAssm        Plots labelling which assumption provides the
                        evidence
plotRetentionArea       Plots the convex increasing retention area
retentionBrd            Computes the Multiplicity Controlled Retention
                        Set.
```

The main function of the evidenceFactors package is retentionBrd.

This package provides tool for reporting of evidence factors analysis in observational studies. Implementation of the functions in this package depends on the computation of maximum p-values for different bias levels for all the evidence factors. Based on the vector of maximum p-values the functions in the package can be used to efficiently perform sensitivity analysis.

The sensitivity analysis asks about the magnitude, gamma, of bias in treatment assignment in observational studies that would need to be present to alter the conclusions of a randomization test that assumed matching for observed covariates removes all bias. The tools and algorithms implemented in evidenceFactors package are discussed in detail in Karmakar et. al. (2016). For general discussion of sensitivity analyses in observational studies, see Chapter 4 of Rosenbaum (2002).

The main function retentionBrd takes as input a list of the maximum p-values for each evidence factors for a range of gamma values. The retention set in such sensitivity analysis is a convex increasing region. The output of the algorithm is the lower boundary of the retention region. The complexity of the algorithm is linear in number of evidence factors and log-linear in the range of

gamma. Either fisher's combination method or the truncated product method (Zaykin et al., 2002) can be used as combining method for evidence factors.

The other two functions in the package are plotting functions. For 2 evidence factors plotRetentionArea produces a grey-scale plot of the retention set, i.e. the set of sensitive bias levels. plotRejDecbyAssm uses a closed testing ideology to produce another grey-scale plot where further shading is used to identify the rejection decisions to one or both of the assumptions. See Karmakar et. al. (2016) for detailed discussion.

Packages sensitivitymv, sensitivitymw and sensitivity2x2xk are various CRAN packages that can be used to compute Plist from dat and Gamlist. See example codes below.

Author(s)

Bikram Karmakar

Maintainer: Bikram Karmakar <bkarmakar@ufl.edu>

References

Karmakar, B., Doubeni, C. A. and Small, D. S. (2020) Evidence factors for in case-control study with application to the effect of flexible sigmiodoscopy screening on colorectal cancer, Annals of Applied Statistics, forthcomming.

Karmakar, B., French, B. and Small, D. S. (2019) Integrating the evidence from evidence factors in observational studies, Biometrika 106 353-367.

Rosenbaum, P. R. (2002) Observational Studies (2nd edition). New York: Springer.

Rosenbaum, P. R. (2010) Evidence factors in observational studies. Biometrika, 97, 333-345.

Rosenbaum, P. R. (2011) Some approximate evidence factors in observational studies. Journal of the American Statistical Association, 106, 285-295.

Zaykin, D., Zhivotovsky, L. A., Westfall, P. and Weir, B. (2002) Truncated product method for combining p-values. Genetic Epidemiology, 22, 170-185.

Examples

```
## mean tail moment data example
library(sensitivitymv)

data(mtm)
Gamseq <- seq(1, 15, by = 0.2)
Gamlist <- list(Gamseq, Gamseq)
Plist <- list(c(), c())

for(gam in Gamseq){
Plist[[1]] = c(Plist[[1]], senmv(-mtm,gamma=gam,trim=1)$pval)
Plist[[2]] = c(Plist[[2]], senmv(-mtm[,2:3],gamma=gam,trim=1)$pval)
}

# Fisher's combination method
rbrd <- retentionBrd(Plist, Gamlist)
plotRetentionArea(rbrd, Gamlist)
plotRejDecbyAssm(rbrd, Gamlist, Plist, alpha = 0.05)
```

```
# truncated product combination
rbrd <- retentionBrd(Plist, Gamlist, method = "Truncated", talpha = .5)
plotRetentionArea(rbrd, Gamlist)
plotRejDecbyAssm(rbrd, Gamlist, Plist, alpha = 0.05)


## Not run:
# Other usages
## Pseudocode for the analysis of the Sigmoidoscopy study in
## Karmakar, Doubeni and Small (2020).
library(haven)

data_reduced = read_dta('matched_cc_analysis.dta')

data_reduced$casetype=NA
data_reduced$casetype[data_reduced$casetype_c== -1]='Controls'
 data_reduced$casetype[data_reduced$casetype_c== 0]='Proximal cancer cases'
 data_reduced$casetype[data_reduced$casetype_c== 1]='Distal cancer cases'
data_reduced$casetype = as.factor(data_reduced$casetype)

data_reduced$fsg <- NA
data_reduced$fsg[data_reduced$fsg_bk==0]='No'
data_reduced$fsg[data_reduced$fsg_bk==1]='Yes'
data_reduced$fsg = as.factor(data_reduced$fsg)

data_reduced$caseflg <- NA
data_reduced$caseflg[data_reduced$caseflg_bk==0]='No'
data_reduced$caseflg[data_reduced$caseflg_bk==1]='Yes'


## Forming the contingency array.
##
stratas <- unique(data_reduced$strata_id)
nstrata = length(stratas)

tab2x2xk_incases <- array(NA, c(2,2,nstrata))
tab2x2xk <- array(NA, c(2,2,nstrata))

dimnames(tab2x2xk) = list(fsg=c('Yes','No'),
casetype=c('Controls', 'Cases'), stratas=stratas)
dimnames(tab2x2xk_incases) = list(fsg=c('Yes','No'),
casetype=c('Proximal', 'Distal'), stratas=stratas)

for(s in stratas){
data.strata = data_reduced[data_reduced$strata_id==s,
c('fsg', 'casetype')]
temptab = table(data.strata$fsg, data.strata$casetype)

tab2x2xk_incases[,,s] = temptab[c('Yes', 'No'),
c('Proximal cancer cases', 'Distal cancer cases')]
```

```
    tab2x2xk[,'Controls',s] = temptab[c('Yes', 'No'), 'Controls']
    tab2x2xk[,'Cases',s] = temptab[c('Yes', 'No'), 'Proximal cancer cases'] +
    temptab[c('Yes', 'No'), 'Distal cancer cases']


    }

    library(sensitivity2x2xk)
    ## calculating the randomized p-values
    mh(tab2x2xk_incases)
    mh(tab2x2xk)

    ## sensitivity analysis
    mh(tab2x2xk_incases, Gamma=1.5)
    mh(tab2x2xk, Gamma=1.3)



    ## End(Not run)



    ## Simulating a case referent study from an infinite cohort.
    ## Replication code for the simualtion results of Section 5.1 of
    ## Karmakar, Doubeni and Small (2020).


    # Treatment assignment in the favourable situation with
    # P(Z = 1) = 1/3

    # treatment effect delta

    # Fix the sample size for the broad cases
    # and the referents nB and nR respectively.


    ##################Finding quantiles
    #### t3 distribution
    # library(AdMit)
    # mit <- list(p = c(2/3, 1/3),
    #             mu = matrix(c(0, .5), 2, 1, byrow = TRUE),
    #             Sigma = matrix(c(1, 1), 2),
    #             df = 3)
    #X <- rMit(1000, mit)
    # quantile(X, .8)

    #### Normal distribution
    # library(ks)
    # X <- rnorm.mixt(1000, mus = c(0, 0.5), sigmas=c(1, 1), props = c(2/3, 1/3))
    # quantile(X, .8)
    #################

    library(sensitivitymv)
```

```
## Not run:
N = 10000

## End(Not run)

effSeq = seq(0, 1, by = .2)
gamSeq = seq(1, 4.5, by =.25)
pZ = 1/3


rejFreqFisher <- array(0, c(length(effSeq), length(gamSeq), length(gamSeq)) )
rejFreqTP <- array(0, c(length(effSeq), length(gamSeq), length(gamSeq)) )


nB = 2000
nR = 2000


for(Itr in 1:N){
for(betaIdx in 1:length(effSeq)){
beta = effSeq[betaIdx]
ZB = rep(NA, nB)
ZR = rep(NA, nR)

ResB = rep(NA, nB)
ResR = rep(NA, nR)

Bdef = 1.17 #1.031
countB = 0
countR = 0

while(countB < nB | countR < nR){
Ztemp =  sample(c(0,1), 1, prob = c((1-pZ), pZ))

# Response Distribution
 Rtemp = rnorm(1) + ifelse(Ztemp==1, beta, 0)
#Rtemp = rt(1, 3)/sqrt(3) + ifelse(Ztemp==1, beta, 0)


if(Rtemp > Bdef){
if(countB < nB){
countB = countB + 1
ZB[countB] = Ztemp
ResB[countB] = Rtemp
}
}

if(Rtemp <= Bdef){
if(countR < nR){
countR = countR + 1
ZR[countR] = Ztemp
```

```
ResR[countR] = Rtemp
}
}
}


## Simulation done

# Lets first consider the p-value based on the Broad Referent comparison
exposedBR = ZB + ZR
BRcase_sizes = 1

# Now the Narrow vs Merginal comparison
# First defining narrow cases

narrow = ResB > median(ResB)
marginal = ResB < median(ResB)

exposedNM = ZB[narrow] + ZB[marginal]
NMcase_sizes = 1
J = 2


for(gam1 in 1:length(gamSeq)){
for(gam2 in 1:length(gamSeq)){
Gam1 = gamSeq[gam1]
Gam2 = gamSeq[gam2]

pNM = Gam1*exposedNM/(Gam1*exposedNM + (J - exposedNM))
evidenceNM = pnorm((sum(ZB[narrow]) -
sum(NMcase_sizes*pNM))/sqrt(sum(NMcase_sizes*pNM*(1-pNM))),
lower.tail=FALSE)

pBR = Gam2*exposedBR/(Gam2*exposedBR + (J - exposedBR))
evidenceBR = pnorm((sum(ZB) -
sum(BRcase_sizes*pBR))/sqrt(sum(BRcase_sizes*pBR*(1-pBR))),
lower.tail=FALSE)

Pvec = c(evidenceNM, evidenceBR)

rejFreqFisher[betaIdx,gam1,gam2]=rejFreqFisher[betaIdx,gam1,gam2]+
( pchisq(-2*log(prod(Pvec)), 2*length(Pvec),
lower.tail=FALSE) < 0.05 )

rejFreqTP[betaIdx , gam1, gam2] <- rejFreqTP[betaIdx, gam1, gam2] +
( truncatedP(Pvec) < 0.05 )
}
}

}
cat(Itr, " ")
#if(!(Itr %% 50)) cat("\n")
}
```

Gamlist   *Gamma list for illustration of the use of the package.*

Description

List of numeric vectors of increasing values. List consists of two numeric vectors. The length of each of these vectors are same as the corresponding vectors in Plist. For the use of Plist and Gamlist see the documentation of the function retentionBrd.

Usage

```
data("Gamlist")
```

See Also

Plist

---

mtm   *DNA damage from exposure to chromium*

Description

The data are from a study by Meibian et al. (2008) concerning possible damage to human DNA from occupational exposure to chromium. There were three matched groups, the control group (cmtm), a low exposure group (e2mtm) and a high exposure group (e1mtm). The exposed individuals all worked at a tannery where chromium was used in the tanning of leather. The highly exposed group (e1) worked at tanning leather. The low exposure group (e2) worked at the same tannery but did not tan leather. The reported values are the mean tail moment (mtm) of the comet assay, a measure of damage to DNA. High values of mtm indicate greater damage.

Usage

```
data(mtm)
```

Format

Each row of mtm is a matched set. The columns refer to the treatment groups mentioned in the description.

Details

These data were used as an example of approximate evidence factors in Rosenbaum (2011). Under the null hypothesis H0 of no treatment effect, there are two approximately independent tests of H0 subject to different biases of nonrandom selection, specifically the comparison of 30 controls and 60 matched tannery workers, and the comparison of 30 low and 30 high exposure tannery workers. The two comparisons may be subjected to sensitivity analyses, say using senmv. The output of senmv for various values of gamma for the two comparisons are used as input in the retentionBrd function. See the documentation for retentionBrd for the example.

Source

Meibian et al. (2008). Used as an example in Rosenbaum (2011).

References

Meibian, Z., Zhijian, C., Qing, C. et al. (2008) Investigating DNA damage in tannery workers occupationally exposed to tivalent chromium using the comet assay. Mutation Research 654, 45-51.

Rosenbaum, P. R. (2011) Some approximate evidence factors in observational studies. Journal of the American Statistical Association, 2011, 106, 285-295.

Examples

```
data(mtm)
```

Plist	*p-value list for illustration of the use of the package.*

Description

List of numeric vectors of increasing values each between 0 and 1.

Usage

```
data("Plist")
```

Details

Each numeric vector is the maximum p-values for the associated gamma values. The source data where these p-values are calculated are not reported in this package. For the use of Plist and Gamlist see the documentation of the function retentionBrd.

See Also

Plist

---

plotRejDecbyAssm   *Plots labelling which assumption provides the evidence*

Description

For two evidence factors this function plots the assumptions for which the hypothesis is rejected in a sensitivity analysis.

Usage

```
plotRejDecbyAssm(retentionBrd, Gamlist, Plist, alpha = 0.05, ...)
```

Arguments

retentionBrd	Output of retentionBrd function for two evidence factors.
Gamlist	List of two numeric vectors of gamma values for two evidence factors.
Plist	A list of two numeric vectors. Each vector are the maximum p-values for evidence factors for sensitivity analysis for increasing gamma values.
alpha	Size of overall type-1 error in analysis. Should be same as in retentionBrd function.
...	Further arguments to be passed to method image and legend used for plotting.

Details

The aim of this function is to assign the rejection decision to one or either of the assumptions for each couple of bias levels.

plotRejDecbyAssm uses a closed testing ideology to produce another grey-scale plot where further shading is used to identify the rejection decisions to one or both of the assumptions. See Karmakar et. al. (2016) for detailed discussion.

Value

A grey-scale plot similar to the output of plotRetentionArea.

Author(s)

Bikram Karmakar

References

Karmakar, B., French, B., Sadakane, A. and Small, D. S. (2016) Integrating the evidence from evidence factors in observational studies.

See Also

plotRetentionArea

Examples

```
data(Plist)
data(Gamlist)
rbrd = retentionBrd(Plist, Gamlist)
plotRejDecbyAssm(rbrd, Gamlist, Plist)

rbrd = retentionBrd(Plist, Gamlist, method = "Truncated")
plotRejDecbyAssm(rbrd, Gamlist, Plist)

## mean tail moment data example
library(sensitivitymv)

data(mtm)
Gamseq <- seq(1, 15, by = 0.2)
Gamlist <- list(Gamseq, Gamseq)
Plist <- list(c(), c())

for(gam in Gamseq){
Plist[[1]] = c(Plist[[1]], senmv(-mtm,gamma=gam,trim=1)$pval)
Plist[[2]] = c(Plist[[2]], senmv(-mtm[,2:3],gamma=gam,trim=1)$pval)
}

# Fisher's combination method
rbrd <- retentionBrd(Plist, Gamlist)
plotRejDecbyAssm(rbrd, Gamlist, Plist)

# truncated product combination
rbrd <- retentionBrd(Plist, Gamlist, method = "Truncated", talpha = .5)
plotRejDecbyAssm(rbrd, Gamlist, Plist)
```

---

plotRetentionArea   *Plots the convex increasing retention area*

Description

For two evidence factors this function produces a grey-scale plot of the retention set, i.e. the set of
sensitive bias levels.

Usage

```
plotRetentionArea(retentionBrd, Gamlist, ...)
```

Arguments

retentionBrd	Output of retentionBrd function for two evidence factors.
Gamlist	List of two vectors of gamma values for two evidence factors.
...	Further arguments to be passed to method image and legend used for plotting.

Details

For 2 evidence factors plotRetentionArea produces a grey-scale plot of the retention set, i.e. the set of sensitive bias levels.

Value

The output of the function is plot.

Author(s)

Bikram Karmakar

References

Karmakar, B., French, B., Sadakane, A. and Small, D. S. (2016) Integrating the evidence from evidence factors in observational studies.

See Also

plotRejDecbyAssm.

Examples

```
data(Plist)
data(Gamlist)
rbrd = retentionBrd(Plist, Gamlist)
plotRetentionArea(rbrd, Gamlist)

rbrd = retentionBrd(Plist, Gamlist, method = "Truncated")
plotRetentionArea(rbrd, Gamlist)

## mean tail moment data example
library(sensitivitymv)

data(mtm)
Gamseq <- seq(1, 15, by = 0.2)
Gamlist <- list(Gamseq, Gamseq)
Plist <- list(c(), c())

for(gam in Gamseq){
Plist[[1]] = c(Plist[[1]], senmv(-mtm,gamma=gam,trim=1)$pval)
Plist[[2]] = c(Plist[[2]], senmv(-mtm[,2:3],gamma=gam,trim=1)$pval)
}

# Fisher's combination method
rbrd <- retentionBrd(Plist, Gamlist)
plotRetentionArea(rbrd, Gamlist)

# truncated product combination
rbrd <- retentionBrd(Plist, Gamlist, method = "Truncated", talpha = .5)
plotRetentionArea(rbrd, Gamlist)
```

---

retentionBrd *Computes the Multiplicity Controlled Retention Set.*

Description

Computes the lower boundary of convex retention set of sensitivity analysis for evidence factors.

Usage

```
retentionBrd(Plist, Gamlist, alpha = 0.05, method = c("Fisher", "TruncatedP"),
  talpha = .2)
```

Arguments

Plist	A list of numeric vectors. Each vector are maximum p-values for evidence factors for sensitivity analysis for increasing gamma values.
Gamlist	A list of numeric vectors. gamma values associated with evidence factors.
alpha	Size of overall type-1 error in analysis.
method	Combination method of evidence factors. Default choice is Fisher's combination. Another option is "TruncatedP" for the truncated product method.
talpha	Level of truncation for truncated product ("TruncatedP") combination method. Default is 0.20.

Details

Vectors in Plist and Gamlist are necessarily increasing.

The reporting of the sensitivity analysis also controlled for overall type-1 error. The retention set in a sensitivity analysis of evidence factors is a convex increasing region. The output of the algorithm is the lower boundary of the retention region. The complexity of the algorithm is linear in number of evidence factors and log-linear in the range of gamma. The pseudo-code for this function with discussion of the complexity of the algorithm is available in Karmakar et. al. (2016).

Two methods of combing methods can be used, "Fisher" and "TruncatedP". For details on the two methods see Karmakar et. al. (2016). The truncated method of combing p-values was first introduced by Zaykin et al. (2002). In practice the truncated product method is recommended even though neither one uniformly dominates the other in terms of power.

Packages sensitivitymv, sensitivitymw and sensitivity2x2xk are various CRAN packages that can be used to compute Plist from dat and Gamlist. See example codes below.

Value

A matrix with number of columns same as the number of evidence factors. This matrix identifies the border of the retention region.

Author(s)

Bikram Karmakar

References

Karmakar, B., French, B., Sadakane, A. and Small, D. S. (2016) Integrating the evidence from evidence factors in observational studies.

Rosenbaum, P. R. (2002) Observational Studies (2nd edition). New York: Springer.

Rosenbaum, P. R. (2010) Evidence factors in observational studies. Biometrika, 97, 333-345.

Rosenbaum, P. R. (2011) Some approximate evidence factors in observational studies. Journal of the American Statistical Association, 106, 285-295.

Zaykin, D., Zhivotovsky, L. A., Westfall, P. and Weir, B. (2002) Truncated product method for combining p-values. Genetic Epidemiology, 22, 170-1

Examples

```
data(Plist)
data(Gamlist)
retentionBrd(Plist, Gamlist)
retentionBrd(Plist, Gamlist, method = "Truncated")

## mean tail moment data example
library(sensitivitymv)

data(mtm)
Gamseq <- seq(1, 15, by = 0.2)
Gamlist <- list(Gamseq, Gamseq)
Plist <- list(c(), c())

for(gam in Gamseq){
Plist[[1]] = c(Plist[[1]], senmv(-mtm,gamma=gam,trim=1)$pval)
Plist[[2]] = c(Plist[[2]], senmv(-mtm[,2:3],gamma=gam,trim=1)$pval)
}

# Fisher's combination method
rbrd <- retentionBrd(Plist, Gamlist)

# truncated product combination
rbrd <- retentionBrd(Plist, Gamlist, method = "Truncated", talpha = .5)
```

Index