Package ‘kin.cohort’

October 13, 2022

Type Package
Title Analysis of Kin-Cohort Studies
Version 0.7
Date 2015-08-15
Author Victor Moreno, Nilanjan Chatterjee, Bhramar Mukherjee
Maintainer Victor Moreno <v.moreno@iconcologia.net>
Depends survival
Description Analysis of kin-cohort studies. kin.cohort provides estimates of age-specific cumulative risk of a disease for carriers and noncarriers of a mutation. The cohorts are retrospectively built from relatives of probands for whom the genotype is known. Currently the method of moments and marginal maximum likelihood are implemented. Confidence intervals are calculated from bootstrap samples. Most of the code is a translation from previous 'MATLAB' code by N. Chatterjee.
License GPL (>= 2)
NeedsCompilation no
Repository CRAN
Date/Publication 2015-08-28 16:36:59

R topics documented:

kc.marginal .......................... 2
kc.moments ................................ 4
kin.cohort .................................. 5
kin.data ................................... 7
methods ..................................... 7
simulations .................................. 9

Index 12
Marginal Maximum Likelihood estimation of kin-cohort data

Description

This function estimates cumulative risk and hazard at given ages for carriers and noncarriers of a mutation based on the probands genotypes. It uses the Marginal Maximum Likelihood estimation method (Chatterjee and Wacholder, 2001). Piece-wise exponential distribution is assumed for the survival function.

Usage

kc.marginal(t, delta, genes, r, knots, f, pw = rep(1,length(t)), set = NULL, B = 1, maxit = 1000, tol = 1e-5, subset, logrank=TRUE, trace=FALSE)

Arguments

t time variable. Usually age at diagnosis or at last follow-up
delta disease status (1: event, 0: no event
genes factor or numeric vector (1 gene), matrix or data.frame (2 genes) with genotypes of proband numeric. factors and data.frame with factors are prefered in order to use user-defined labels. Otherwise use codes (1:noncarrier, 2: carrier, 3: homozygous carrier)
r relationship with proband 1:parent, 2:sibling 3:offspring 0:proband. Probands will be excluded from analysis and offspring will be recoded 1 internally.
knots time points (ages) for cumulative risk and hazard estimates
f vector of mutation allele frequencies in the population
pw prior weights, if needed
set family id (only needed for bootstrap)
B number of bootstrap samples (only needed for bootstrap)
maxit max number of iterations for the EM algorithm
tol convergence tolerance
subset logical condition to subset data
logrank Perform a logrank test
trace Show iterations for bootstrap

Value

object of classes "kin.cohort" and "chatterjee".
cumrisk matrix with cumulative risk estimates for noncarriers, carriers and the cumulative risk ratio. Estimates are given for the times indicated in the knot vector
**hazard**
matrix with hazard estimates for noncarriers, carriers and the hazard ratio. Estimates are given for the times indicated in the knot vector

**knots**
vector of knots

**conv**
if the EM algorithm converged

**niter**
number of iterations needed for convergence

**ngen.rel**
number of combinations of genotypes in the relatives

**events**
matrix with number of events and person years per each knot

**logHR**
mean log hazard ratio estimate (unweighted)

**logrank**
logrank test. If 2 genes, for the main effects, the cross-classification and the stratified tests

**call**
copy of call

if bootstrap confidence intervals are requested (B>1) then the returned object is of classes "kin.cohort.boot" and "chatterjee" with previous items packed in value estimate and each bootstrap sample packed in matrices.

**Note**
This function is best called by kin.cohort than directly

**References**

**See Also**
k.in.cohort, print.kin.cohort, plot.kin.cohort

**Examples**
```r
## Not run:
data(kin.data)
attach(kin.data)
res.mml<- kc.marginal(age, cancer, gen1, rel, knots=c(30,40,50,60,70,80), f=0.02)
res.mml
## End(Not run)
```
**kc.moments**

**Kin-cohort estimation of penetrance by the method of moments**

**Description**

This function estimates cumulative risk and hazard at given ages for carriers and noncarriers of a mutation based on the probands genotypes. It uses the method of moments described by Wacholder et al (1998)

**Usage**

```
kc.moments(t, delta, genes, r, knots, f, pw = rep(1,length(t)),
           set = NULL, B = 1, logrank = TRUE, subset, trace=FALSE)
```

**Arguments**

- `t`: time variable. Usually age at diagnosis or at last follow-up
- `delta`: disease status (1: event, 0: no event
- `genes`: genotype of proband numeric. A factor is preferred, otherwise numeric code of genotypes (1: noncarrier, 2:carrier, [3: homozygous carrier])
- `r`: relationship with proband 1:parent, 2:sibling 3:offspring 0:proband. Probands will be excluded from analysis and offspring will be recoded 1 internally.
- `knots`: time points (ages) for cumulative risk and hazard estimates
- `f`: mutation allele frequency in the population
- `pw`: prior weights, if needed
- `set`: family id (only needed for bootstrap)
- `B`: number of bootstrap samples (only needed for bootstrap)
- `logrank`: if logrank test is desired
- `subset`: logical condition to subset data
- `trace`: Show iterations for bootstrap

**Value**

object of classes "kin.cohort" and "wacholder".

- `cumrisk`: matrix of dimension (number of knots x 3) with cumulative risk festimates or noncarriers, carriers and the cumulative risk ratio
- `knots`: vector of knots
- `km`: object class survfit (package survival)
- `logrank`: p-value of the logrank test
- `events`: matrix with number of events and person years per each knot
- `call`: copy of call

if bootstrap confidence intervals are requested (B>1) then the returned object is of classes "kin.cohort.boot" and "wacholder" with previous items packed in value estimate and each bootstrap sample packed in matrices.
kin.cohort

Note
This function is best called by kin.cohort than directly

References

See Also
kin.cohort, print.kin.cohort, plot.kin.cohort

Examples
## Not run:
data(kin.data)
attach(kin.data)
res.km<- kc.moments(age, cancer, gen1, rel, knots=c(30,40,50,60,70,80), f=0.02)
res.km
## End(Not run)

kin.cohort
Analysis of kin-cohort data

Description
This function estimates cumulative risk at given ages for carriers and noncarriers of a mutation based on the probands genotypes. It can use the Marginal Maximum Likelihood estimation method (Chatterjee and Wacholder, 2001) or the method of moments (Wacholder et al, 2001). Bootstrap confidence intervals can be requested.

Usage
kin.cohort(..., method = c("marginal", "mml", "chatterjee", "moments", "km", "watcholder"))

Arguments
... see kc.marginal and kc.moments for details
method choose estimation method: Marginal Maximum Likelihood (selected by "marginal", "mml", "chatterjee") or method of moments (selected by "moments", "km", "watchholder")

Details
This function is a wrapper that will call kc.marginal or kc.moments depending on the argument method.
Author(s)
Victor Moreno, Nilanjan Chatterjee, Bhramar Mukherjee.
Maintainer: Victor Moreno <v.moreno@iconcologia.net>

References

See Also
kc.marginal, kc.moments

Examples
```r
## Not run:
data(kin.data)
attach(kin.data)

res.k<- kin.cohort(age, cancer, gen1, rel, knots=c(30,40,50,60,70,80), f=0.02, method="km")
res.k
plot(res.k)
plot(res.k,what="crr")

set.seed(1)
res.k.b<- kin.cohort(age, cancer, gen1, rel, knots=c(30,40,50,60,70,80), f=0.02, set=family, method="km", B=10)
res.k.b
plot(res.k.b)
plot(res.k.b,what="crr")

res.m<- kin.cohort(age, cancer, gen1, rel, knots=c(30,40,50,60,70,80), f=0.02, method="mml")
res.m
plot(res.m)
plot(res.m, what="hr")

res.m2<- kin.cohort(age, cancer, data.frame(gen1,gen2), rel, knots=c(30,40,50,60,70,80), f=c(0.02,0.01), method="mml")
res.m2
plot(res.m2)
plot(res.m2, what="hr")

set.seed(1)
res.m.b<- kin.cohort(age, cancer, gen1, rel, knots=c(30,40,50,60,70,80), f=0.02, set=family, method="mml", B=10)
res.m.b
plot(res.m.b)
```
kin.data

plot(res.m.b, what="hr")

## End(Not run)

---

**kin.data**

*sample data for kin-cohort analysis*

---

**Description**

Simulated data of a study on the penetrance of breast cancer for carriers 2 mutations.

**Usage**

data(kin.data)

**Format**

A data frame with 15341 observations on the following 5 variables.

- **age**  age at diagnosis or at last follow-up
- **cancer**  disease status (1: breast cancer, 0: no breast cancer
- **gen1**  gen1 genotypes of proband
- **gen2**  gen2 genotypes of proband
- **rel**  relationship with proband 1:parent or offspring, 2:sibling
- **family**  family id

**Examples**

data(kin.data)

---

**methods**

*methods for print and plot*

---

**Description**

Functions to print a formatted output and produce plots
Usage

## S3 method for class 'kin.cohort'
print(x, descriptive = TRUE, cumrisk = TRUE, hazard = FALSE, survival = FALSE, logrank = TRUE, HR = TRUE, digits = 5, ...)

## S3 method for class 'kin.cohort.boot'
print(x, cumrisk = TRUE, hazard = FALSE, HR = TRUE, conf = 0.95, digits = 5, show = TRUE, logrank = TRUE, ...)

## S3 method for class 'kin.cohort'
plot(x, what = c("cr", "hr", "crr"), min.age = min(x$knots), max.age = max(x$knots), max.y, type, add = FALSE, color, line, ...)

## S3 method for class 'kin.cohort.boot'
plot(x, conf = 0.95, what = c("cr", "hr", "crr"), min.age = min(x$knots), max.age = max(x$knots), age.start = 0, start.ref, max.y, type, median = FALSE, add = FALSE, color, line, ...)

Arguments

x  object to be printed or plotted
descriptive  print table with number of events and person-years
cumrisk  print cumulative risk
hazard  print hazard
survival  print survival
HR  print harad ratios
logrank  print logrank p value
digits  digits for rounding
show  do not print
conf  coverage for confidence intervals
what  type of plot desired: cumulative risk ("cr"), hazard ratio ("hr", for marginal method only), cumulative risk ratio ("crr", for moments method only)
min.age  Minimal age for plots
max.age  Maximal age for plots
age.start  initial age value (x) for plots
start.ref  initial risk value (y) for plots
max.y  Max value for y axis
type  type of line in plots
add  If TRUE, then lines are added to current plot. Useful to compare analyses.
color  change line colors using a vector of values
line  change line width using a vector of values
median  plot median of bootstrap samples instead of point estimates
...  additional arguments for print or plot
Details

Specific output and plot types can be selected with arguments

Description

Functions to simulate data for kin-cohort analysis

Usage

kc.simul(nfam, f, hr, rand = 0, mean.sibs = 2, mean.desc = 1.5,
         a.age = 8, b.age = 80, a.cancer = 3, b.cancer = 180)

sample.caco(object, p.cases = 1, caco.ratio = 1, verbose = TRUE)

## S3 method for class 'kin.cohort.sample'
summary(object,...)

Arguments

nfam  number of families to be generated
f   allele frequency
hr  hazard ratio for disease carriers relative noncarriers
rand variance of random effect for cancer incidence (ratio of hr)
mean.sibs mean number of sibllings and descendants (~Poisson)
mean.desc mean number of sibllings and descendants (~Poisson)
a.age shape parameter for age (~Weibull)
b.age scale parameter for age (~Weibull)
a.cancer shape parameter for cancer incidence (~Weibull)
b.cancer scale parameter for cancer incidence (~Weibull)
object object of class kin.cohort.sample and data.frame
p.cases proportion of cases (affected) to include in sample. if more than 1, the exact
number is assumed
caco.ratio ratio of controls per case to include in sample
verbose show the number of cases and controls sampled
... additional arguments
Details

kc.simul will generate a cohort of probands of size nfam. Default parameters simulate a typical cancer study. Each of them will be assigned: a carrier status with probability \( f^2 + 2f(1 - f) \); a current age drawn from a Weibull distribution with parameters a. age and b. age; an age at diagnosis (agecancer) drawn from a Weibull distribution with parameters a. cancer and b. cancer, if noncarrier. For carriers, the scale (b. cancer) is shifted to get the desired hazard ratio (hr). If rand>0, then a family specific random effect is also added, drawn from a normal distribution with mean 0 and sd rand. If agecancer< age then the disease status (cancer) will be 1, 0 otherwise.

First degree relatives are generated for each proband: two parents, a random number of sibblings (drawn from a Poisson with the mean mean.sibs), and a random number of descendants (drawn from a Poisson with mean mean.desc). Each of them is assigned a carrier status with probability according to mendelian transmission conditional of the proband carrier status. Current age for relatives are generated conditional on the proband’s age, with random draws from normal distribution. Age at diagnosis (agecancer) is assumed independent, except for the optional family random effect. Gender is assigned at random with probability 0.5 for all individuals.

Note that the simulation of residual familial correlation with a random effect (rand$>0$) does not maintain the desired hazard ratio (hr).

The generic function summary will show the number and proportion of carriers and affected subjects in the sample.

sample.caco will sample (from a simulation generated by kc.simul) a subset of cases (affected probands) and controls (unaffected probands) and their relatives. Currently only random sampling of controls is implemented (no matching). Sampling fraction is controled by caco.ratio.

Currently, only one gene and one disease are simulated.

Value

object of class kin.cohort.sample and data.frame with fields

- famid: family id
- rel: relative type (0=proband, 1=parents, 2=sibblings, 3=descendants)
- age: current age of each subject
- gender: gender (0=male, 1=female)
- carrier: carrier status of proband (0=noncarrier, 1=carrier), common for all family members
- cancer: affected (0=no, 1=yes)
- agecancer: age at diagnosis or current age if not affected
- real.carrier: carrier status or relatives (0=noncarrier, 1=carrier )

Examples

```r
## Not run:
set.seed(7)
## cohort
s<-kc.simul(4000, f=0.03, hr=5)
summary(s)
```
## exclude probands
m.coh<- kc.marginal(s$agecancer, s$cancer, factor(s$carrier), s$rel,
   knots=c(30,40,50,60,70,80,90), f=0.03)
m.coh

## relatives only
r.coh<- coxph(Surv(agecancer,cancer)~real.carrier, data=s)
print(exp(coef(r.coh)))

## probands only
p.coh<- coxph(Surv(agecancer,cancer)~carrier, data=s)
print(exp(coef(p.coh)))

## case-control
s.cc<- sample.caco(s)
summary(s.cc)

## exclude probands
m.caco<- kc.marginal(s.cc$agecancer, s.cc$cancer, factor(s.cc$carrier),
   s.cc$rel, knots=c(30,40,50,60,70,80,90), f=0.03)
m.caco

## relatives only
r.caco<- glm(cancer~real.carrier, family=binomial, data=s.cc, subset=(s.cc$rel!=0))
print(exp(coef(r.caco)[2]))

## probands only
p.caco<- glm(cancer~carrier, family=binomial, data=s.cc, subset=(s.cc$rel==0))
print(exp(coef(p.caco)[2]))

## End(Not run)
Index

* datasets
  kin.data, 7

* methods
  methods, 7
  simulations, 9

* print
  methods, 7

* survival
  kc.marginal, 2
  kc.moments, 4
  kin.cohort, 5
  simulations, 9

  kc.marginal, 2, 5, 6
  kc.moments, 4, 5, 6
  kc.simul(simulations), 9
  kin.cohort, 3, 5, 5
  kin.data, 7

  methods, 7

  plot.kin.cohort, 3, 5
  plot.kin.cohort(methods), 7
  print.kin.cohort, 3, 5
  print.kin.cohort(methods), 7

  sample.caco(simulations), 9
  simulations, 9
  summary.kin.cohort.sample
    (simulations), 9