

Package ‘cmprskContin’

November 15, 2009

Type Package

Title Continuous mark-specific relative risks for two groups

Version 1.5

Date 2009-11-12

Author Peter Gilbert, Ian McKeague, and Yanqing Sun

SystemRequirements Linux/64bit

Maintainer Valerie Obenchain <vobencha@fhcrc.org>

Description Estimation and testing of continuous mark-specific relative risks in two groups as described in Gilbert, P., McKeague, I. and Sun, Y. (2008) The 2-sample problem for failure rates depending on a continuous mark: an application to vaccine efficacy. *Biostatistics* 9, 2, 263-276. This package implements the methods presented in the paper for testing mark-specific hazards ratios and for estimation of mark-specific incidence ratios that are cumulative in time or cumulative in both time and the continuous mark.

License GPL-2

Repository CRAN

LazyLoad yes

OS_type unix

Date/Publication 2009-11-15 13:51:08

R topics documented:

cmprskContin-package	2
cmprsk	3
plotCmprsk	8
summaryCmprsk	10

Index	13
--------------	-----------

cmprskContin-package

continuous mark-specific relative risks for two treatment groups

Description

Estimation and testing of continuous mark-specific relative risks in two groups as described in Gilbert, McKeague and Sun (2008). Results include point and confidence interval estimates, test statistics, and corresponding p-values.

Details

Package: cmprskContin
Type: Package
Version: 1.5
Date: 2009-11-12
License: GPL-2
LazyLoad: yes

Motivations for this work were applications in preventive HIV vaccine trials and therefore results are returned as vaccine efficacies (VEs) as well as relative risks (RRs) where $RR = 1 - VE$. In the following equations, λ_k represents the mark-specific hazard rate for group k , $k=1,2$. Vaccine efficacy is defined as the relative reduction in infection hazard due to vaccination, $VE(t, v) = 1 - \frac{\lambda_1(t, v)}{\lambda_2(t, v)}$ where t is the time of infection and v is the mark (e.g., the genetic distance of an infecting virus from the vaccine strain).

Tests are provided for $H_0^0 : \lambda_1(t, v) = \lambda_2(t, v)$, against the following alternative hypotheses:

$$H_1^0 : \lambda_1(t, v) \leq \lambda_2(t, v)$$

$$H_2^0 : \lambda_1(t, v) \neq \lambda_2(t, v)$$

If H_0^0 is rejected, then it is of interest to assess if the relative risk varies with the mark. For this question, tests are provided for $H_0 : \frac{\lambda_1(t, v)}{\lambda_2(t, v)}$ does not depend on v for $t \in [0, \tau]$, against the following alternative hypotheses:

$$H_1 : \frac{\lambda_1(t, v_1)}{\lambda_2(t, v_1)} \leq \frac{\lambda_1(t, v_2)}{\lambda_2(t, v_2)}$$

$$H_2 : \frac{\lambda_1(t, v_1)}{\lambda_2(t, v_1)} \neq \frac{\lambda_1(t, v_2)}{\lambda_2(t, v_2)}$$

Both integration-type and supremum-type statistics are provided for testing the hypotheses. Test statistics U_1^1 and U_2^1 provide evidence against H_0^0 in the direction of H_1^0 and U_3^1 and U_4^1 provide evidence against H_0^0 in the direction of H_2^0 . For r as np (nonparametric) or sp (semiparametric), U_1^r and U_2^r measure departures from H_0 in the direction of H_1 and H_2 , respectively.

P-values are computed using a Gaussian multipliers simulation procedure.

Results provided include the cumulative incidence estimate of overall VE, proportional hazards estimate of overall VE, log-rank statistic for comparing overall hazard rates, test statistics for evaluating the hypotheses outlined above, and the corresponding p-values.

There are 3 functions :

```
cmprsk
plotCmprsk
summaryCmprsk
```

Author(s)

Peter Gilbert, Ian McKeague, and Yanqing Sun
 Maintainer: Valerie Obenchain <vobencha@fhcrc.org>

References

Gilbert, P., McKeague, I. and Sun, Y. (2008) *The 2-sample problem for failure rates depending on a continuous mark: an application to vaccine efficacy*. *Biostatistics* 9, 2, 263-276.
 Andersen, P. K., Borgan, O., Gill, R. D., and Keiding. (1993) *Statistical Models Based on Counting Processes*. New York: Springer.

See Also

cmprsk, summaryCmprsk and plotCmprsk functions

cmprsk

continuous mark-specific relative risks for two treatment groups

Description

Estimation and testing of continuous mark-specific relative risks in two groups as described in Gilbert, McKeague and Sun (2008, *Biostatistics*). Motivations for this work were applications in preventive HIV vaccine efficacy trials and therefore results are returned as mark-specific vaccine efficacies (VEs) as well as mark-specific relative risks (RRs) where $RR = 1 - VE$. This function implements the paper's methods for testing of mark-specific hazards ratios (i.e., relative risks), and for estimation of mark-specific cumulative incidence ratios (cumulative in time) and of mark-specific doubly cumulative incidence ratios (cumulative in both time and the continuous mark).

Usage

```
cmprsk(gp, ftime, ftype, mark, nboot=5000, ngrid, ngridv,
T1=0, T2=0, UT1=0, UT2=0, ttanal=0, BAND1=0, BAND2=0, TAILSL=1,
TAILSU=1, V1=0, V2=0, UV1=0, UV2=0, BANDV1=0, BANDV2=0, BANDVLOW=0,
BANDVUP=0, TAILSV=1)
```

Arguments

gp	Vector of 0 and 1 indicating subject group. Missing values are not allowed.
ftime	Vector of right-censored failure times. This value should be the earliest of event time (i.e., HIV infection) and censoring time. Missing values are not allowed.
ftype	Vector indicating 1=failure (infected) or 0=right censored (uninfected). Missing values are not allowed.
mark	Vector of continuous mark value. This value should be in [0,1] if infected and measured, 99 otherwise. The mark for all subjects with <code>ftype=0</code> should be set to 99 (i.e., mark does not exist for these subjects). A mark value of 99 should be used for subjects with <code>ftype=1</code> if the mark is missing or was not measured. This function performs a complete-case analysis and the <code>ftype=1</code> subjects with a mark value of 99 will be excluded.
nboot	Number of bootstrap replicates for computing p-values, <code>nboot ≥ 5000</code> is recommended
ngrid	(optional) Number of failure time gridpoints for computing the test process. If specified, this value must be less than the length of the <code>ftime</code> vector. Default is <code>round((0.5)*length(ftime))</code> .
ngridv	(optional) Number of mark gridpoints for computing the test process. If specified, this value must be less than the number of subjects with <code>ftype=1</code> . Default is <code>round((0.5)*length(which(ftype==1)))</code> .
T1	(optional) Smallest failure time used to compute the test statistics. This value should be just larger than the smallest observed failure time. The default (indicated by <code>T1=0</code>) uses the smallest failure time such that there are 2 failure times smaller than T1 in each group.
T2	(optional) Largest failure time used to compute the test statistics. This value should be just smaller than the largest observed failure time. The default (indicated by <code>T2=0</code>) is the largest failure time such that there are 2 failure times larger than T2 in each group.
UT1	(optional) Lower envelope for smoothing the failure time ($UT1 \leq T1$). The default is $UT1 = \frac{T1}{2}$.
UT2	(optional) Upper envelope for smoothing the failure time ($T2 \leq UT2$). It is recommended to set <code>UT2</code> equal to the largest observed failure time or a bit bigger. The default (indicated by <code>UT2=0</code>) uses the smallest of the largest right-censored failure times for the two groups.
ttanal	(optional) The failure time (t) at which $VE^c(t, v)$ and $VE^{dc}(t, v)$ are evaluated. The default (indicated by <code>ttanal=0</code>) is <code>ttanal=T2</code> .
BAND1	(optional) Bandwidth for kernel estimation of the group 1 hazard over time. The default (indicated by <code>BAND1=0</code>) is the optimal bandwidth, defined as the bandwidth that minimizes an asymptotic approximation of the mean integrated squared error of $\hat{\lambda}_1(t)$ with the second derivative bandwidth set equal to $\frac{(T2-T1)}{4}$. Details of this optimization are in Andersen, P. K., Borgan, O., Gill, R. D., and Keiding (1993).
BAND2	Bandwidth for kernel estimation of the group 2 hazard over time. The default (indicated by <code>BAND2=0</code>) is the optimal bandwidth chosen the same as for group1.

TAILSL	(optional) Left tail-correction for kernel estimation of the hazard functions over time. 1=use correction (default), 2=no correction.
TAILSU	(optional) Right tail-correction for kernel estimation of the hazard functions over time. 1=use correction (default), 2=no correction.
V1	(optional) Left limit for smoothing in the mark (V). The default (indicated by $V1=0$) is the largest of the smallest observed marks for the 2 groups.
V2	(optional) Right limit for smoothing in the mark (V). The default (indicated by $V2=0$) is the smallest of the of the largest observed marks for the 2 groups.
UV1	(optional) Left envelope for smoothing in the mark; must have $UV1 \leq V1$. The default (indicated by $UV1=0$) is $UV1 = V1$.
UV2	(optional) Right envelope for smoothing in the mark; must have $V2 \leq UV2$. The default (indicated by $UV2=0$) is $UV2 = V2$.
BANDV1	(optional) Bandwidth for smoothing in the group 1 mark. The default (indicated by $BANDV1=0$) is to use the optimal bandwidth, chosen to minimize the 2-fold cross-validated mean integrated squared error.
BANDV2	(optional) Bandwidth for smoothing in the group 2 mark. The default (indicated by $BANDV2=0$) is to use the optimal bandwidth, chosen to minimize the 2-fold cross-validated mean integrated squared error.
BANDVLOW	(optional) The lower range of the mark values to search over for computing the optimized $BANDV1$ and $BANDV2$ arguments. The default (indicated by $BANDVLOW=0$) is 5% of the way from the maximum smallest observed mark in the 2 groups to the $BANDVUP$ value.
BANDVUP	(optional) The upper range of mark values to search over to compute the optimized $BANDV1$ and $BANDV2$ arguments. The default (indicated by $BANDVUP=0$) is the maximum mark divided by 2.
TAILSV	(optional) Upper and lower tail-correction for smoothing in the mark. 1=smoothing (default), 2=no smoothing.

Details

This package performs a complete case analysis, restricting to subjects with complete data. Missing values for `gp`, `ftime` and `ftype` are not allowed. For all subjects with `ftype=0` (uninfected), the mark value should be set to 99. Subjects with `ftype=1` (infected) should have a mark value in $[0,1]$. If the mark is missing or was not measured for these `ftype=1` subjects, the mark should be set to 99 and they will be excluded from the analysis.

The results are returned as a list. The first list item, `STATvec` contains test statistics, p-values and the estimated $VE(t, v)$ or $RR(t, v)$ over the grid of values for v , at the fixed time-point `ttanal`. The second and third items, `VECmat` and `VEDCmat`, are matrices containing point and 95% confidence interval estimates of the cumulative and doubly cumulative $VE(t, v)$ or $RR(t, v)$ over the grid of values for v , respectively, at time-point `ttanal`. The fourth item, `timeMark`, is a matrix containing the point and confidence interval estimates of the cumulative and doubly cumulative $VE(t, v)$ or $RR(t, v)$, over the grid of time-points t between `T1` and `T2` as well as over the grid of values for v .

Value

A 4-item list is returned, `STATvec`, `VECmat`, `VEDCmat`, and `timeMark`. The components of each list are detailed below.

<code>STATvec</code>	A vector containing test statistics and p-values.
<code>nsamp1</code>	Number of group 1 subjects
<code>nsamp2</code>	Number of group 2 subjects
<code>nboot</code>	Number of bootstrap replicates used for computing p-values
<code>T1</code>	Smallest failure time used to compute the test statistics
<code>T2</code>	Largest failure time used to compute the test statistics
<code>ttanal</code>	The failure time (t) at which $VE^c(t, v)$ and $VE^{dc}(t, v)$ are evaluated
<code>AvgV1</code>	Average mark among infected group 1 subjects
<code>AvgV2</code>	Average mark among infected group 2 subjects
<code>BAND1</code>	Bandwidth for estimating the group 1 hazard over time
<code>BAND2</code>	Bandwidth for estimating the group 2 hazard over time
<code>BANDV1</code>	Bandwidth for smoothing in the group 1 mark
<code>BANDV2</code>	Bandwidth for smoothing in the group 2 mark
<code>V1</code>	The smallest mark used for estimation (used for plotting)
<code>V2</code>	The largest mark used for estimation (used for plotting)
<code>VEhatCI</code>	Cumulative incidence estimate of overall VE
<code>VEhatPH</code>	Proportional hazard estimate of overall VE
<code>LogRankZ</code>	Log-rank statistic for comparing overall hazard rates
<code>U11</code>	Nonparametric test statistic for 1-sided testing of $H_0^0 : VE(t, v) = 0$
<code>U12</code>	Nonparametric test statistic for 1-sided testing of $H_0^0 : VE(t, v) = 0$
<code>U13</code>	Nonparametric test statistic for 2-sided testing of $H_0^0 : VE(t, v) = 0$
<code>U14</code>	Nonparametric test statistic for 2-sided testing of $H_0^0 : VE(t, v) = 0$
<code>pval11</code>	p-value corresponding to U11
<code>pval12</code>	p-value corresponding to U12
<code>pval13</code>	p-value corresponding to U13
<code>pval14</code>	p-value corresponding to U14
<code>Unp1</code>	Nonparametric test statistic for 1-sided testing of $H_0 : VE(t, v) = VE(t)$
<code>Unp2</code>	Nonparametric test statistic for 2-sided testing of $H_0 : VE(t, v) = VE(t)$
<code>pvalnp1</code>	p-value corresponding to Unp1
<code>pvalnp2</code>	p-value corresponding to Unp2
<code>Usp1</code>	Semiparametric test statistic for 1-sided testing of $H_0 : VE(t, v) = VE(t)$
<code>Usp2</code>	Semiparametric test statistic for 2-sided testing of $H_0 : VE(t, v) = VE(t)$
<code>pvalsp1</code>	p-value corresponding to Usp1
<code>pvalsp2</code>	p-value corresponding to Usp2
<code>VECmat</code>	A <code>ngridv</code> x 11 matrix containing data for estimating $VE^c(ttanal, v)$ vs v .
<code>index</code>	Index of the analysis time <code>ttanal</code>
<code>ttanal</code>	Failure time at which the cumulative VE is evaluated
<code>mark</code>	Mark value
<code>F1</code>	Estimated cumulative incidence function $F_1(ttanal, v)$ for group 1
<code>F2</code>	Estimated cumulative incidence function $F_2(ttanal, v)$ for group 2

varF1	Estimated variance of F1
varF2	Estimated variance of F2
VEC	Estimated cumulative vaccine efficacy, $1 - \frac{F_1(ttanal, v)}{F_2(ttanal, v)}$
varVEC	Estimated variance of VEC
CIlow	Lower 95% CI limit for VEC
CIhigh	Upper 95% CI limit for VEC
VEDCmat	An ngridv x 11 matrix containing data for estimating $VE^{dc}(ttanal, v)$ vs v .
index	Index of the analysis time $ttanal$
ttanal	Failure time at which the doubly cumulative VE is evaluated
mark	Mark value
F1dc	Estimated doubly cumulative incidence function for group 1
F2dc	Estimated doubly cumulative incidence function for group 2
varF1dc	Estimated variance of doubly cumulative incidence function for group 1
varF2dc	Estimated variance of doubly cumulative incidence function for group 2
VEDC	Estimated doubly cumulative vaccine efficacy
varVEDC	Estimated variance of VEDC
CIlow	Lower 95% CI limit for VEDC
CIhigh	Upper 95% CI limit for VEDC
timeMark	A matrix containing data for estimating $VE^c(t, v)$ vs (t, v) at all event times t between T1 and T2 (ie, all times between the smallest and largest failure times used for computing the test statistics).
nsamp	Sample index
eventtime	Observed event times
mark	Mark value
F1	Estimated cumulative incidence function $F_1(t, v)$ for group 1
F2	Estimated cumulative incidence function $F_2(t, v)$ for group 2
varF1	Estimated variance of F1
varF2	Estimated variance of F2
VEC	Estimated cumulative vaccine efficacy
varVEC	Estimated variance of cumulative vaccine efficacy

Author(s)

Peter Gilbert, Ian McKeague, and Yanqing Sun
 Maintainer: Valerie Obenchain <vobencha@fhcrc.org>

References

Gilbert, P., McKeague, I. and Sun, Y. (2008) *The 2-sample problem for failure rates depending on a continuous mark: an application to vaccine efficacy*. *Biostatistics* 9, 2, 263-276.
 Andersen, P. K., Borgan, O., Gill, R. D., and Keiding. (1993) *Statistical Models Based on Counting Processes*. New York: Springer.

See Also

summaryCmprsk and plotCmprsk functions

Examples

```
## Simulate data
set.seed(100)
gp <- rep(0:1, times=50)
ftime <- 50*runif(100)
ftype <- rep(0:1, each=50)
# uninfected subjects do not have mark value; use 99
infected <- which(ftype==1)
mark <- rep(99, length(gp))
# create mark values for the infected
mark[infected] <- runif(length(infected))
# create some infected subjects with missing mark; use 99
missing <- sample(infected, 2)
mark[missing] <- 99

## Compute risk
result <- cmprsk(gp, ftime, ftype, mark, nboot=500, T1=2, T2=36, UT1=1,
UT2=42, ttanal=36)

## See cumulative incidence and proportional hazard estimates of overall VE
result[[1]]$VEhatCI
result[[1]]$VEhatPH
```

plotCmprsk

plot estimated vaccine efficacy or relative risk

Description

This function plots the cumulative or doubly cumulative estimated vaccine efficacy $VE(ttanal, v)$ or relative risk $RR(ttanal, v)$ versus the mark v . 95% confidence interval estimates versus the mark v are also plotted. All plots are output to .ps files.

Usage

```
plotCmprsk(x, plottype=1, dc=TRUE, p1, p2, filename, main=" ",
ylim=c(-1,1.5), xlim, legloc=1, xlab="Mark",
ylab="Estimated Risk", ...)
```

Arguments

x	List output from cmprsk function
plottype	Binary 0 or 1 to indicate plot type. 1=vaccine efficacy (default), 0=relative risk

dc	Logical to indicate VE type. TRUE=doubly cumulative VE (default), FALSE=cumulative VE
p1	P-value for testing $H_0^0 : VE(t, v) = 0$ or $H_0^0 : RR(t, v) = 1$ for all times (t) and marks (v). Inputs can be pval11, pval12, pval13, or pval14 from the STATvec component of the cmprsk output list. These p-values are for one-sided hypothesis tests. The default is pval13
p2	P-value for testing $H^0 : VE(t, v) = VE(t)$ or $H^0 : RR(t, v) = RR(t)$ for all times (t) and marks (v). Inputs can be pvalnp1 or pvalnp2 from the STATvec component of the cmprsk output list. These p-values are for two-sided hypothesis tests. The default is pvalnp2.
filename	String name for the output plot file. All files are output as .ps files, the string does not need to include the .ps extension. This is a required argument.
main	Main title for the plot
ylim	y axis limits for plot
xlim	x axis limits for plot. If no xlim is supplied, the default used is xlim=(V1,V2).
legloc	Vector of length 2 specifying the upper right coordinates of the legend; otherwise the legend is placed in the upper right corner of the plot
xlab	x axis label
ylab	y axis label
...	Additional graphical parameters to pass to plot function

Details

A labeled plot of cumulative or doubly cumulative VE or RR plot with confidence intervals and p-values is produced. Only data with mark values between the smallest and largest mark values used in the estimation are plotted.

Author(s)

Peter Gilbert, Ian McKeague, and Yanqing Sun
 Maintainer: Valerie Obenchain <vobencha@fhcrc.org>

References

Gilbert, P., McKeague, I. and Sun, Y. (2008) *The 2-sample problem for failure rates depending on a continuous mark: an application to vaccine efficacy*. *Biostatistics* 9, 2, 263-276.
 Andersen, P. K., Borgan, O., Gill, R. D., and Keiding. (1993) *Statistical Models Based on Counting Processes*. New York: Springer.

See Also

summaryCmprsk and cmprsk functions

Examples

```

## Simulate data
set.seed(100)
gp <- rep(0:1, times=50)
ftime <- 50*runif(100)
ftype <- rep(0:1, each=50)
# uninfected subjects do not have mark value; use 99
infected <- which(ftype==1)
mark <- rep(99, length(gp))
# create mark values for the infected
mark[infected] <- runif(length(infected))
# create some infected subjects with missing mark; use 99
missing <- sample(infected, 2)
mark[missing] <- 99

## Compute risk
x <- cmprsk(gp, ftime, ftype, mark, nboot=500, T1=2, T2=36, UT1=1, UT2=42, ttanal=36)

## Plots
V1 <- x[[1]]$V1
V2 <- x[[1]]$V2
ttanal <- x[[1]]$ttanal
xlab <- "Strain Distance v"

# VE doubly cumulative
plotCmprsk(x, plottype=1, dc=TRUE, filename="plotVEdc",
main=paste("Estimated VE^dc(t,v) vs Distance v, at time-point t=",
ttanal, sep=""), ylim=c(-1, 1.5), xlim=c(V1, V2), legloc=1,
xlab=xlab, ylab="Estimated VE^dc(t,v)")

# VE cumulative
plotCmprsk(x, plottype=1, dc=FALSE, filename="plotVEc",
main=paste("Estimated VE^c(t,v) vs Distance v, at time-point t=",
ttanal, sep=""), ylim=c(-1, 1.5), xlim=c(V1, V2), legloc=1,
xlab=xlab, ylab="Estimated VE^c(t,v)")

# RR doubly cumulative
plotCmprsk(x, plottype=2, dc=TRUE, filename="plotRRdc",
main=paste("Estimated RR^dc(t,v) vs Distance v, at time-point t=",
ttanal, sep=""), ylim=c(0.0, 4.5), xlim=c(V1, V2), legloc=1,
xlab=xlab, ylab="Estimated RR^dc(t,v)")

# RR cumulative
plotCmprsk(x, plottype=2, dc=FALSE, filename="plotRRc",
main=paste("Estimated RR^c(t,v) vs Distance v, at time-point t=",
ttanal, sep=""), ylim=c(0.0, 4.5), xlim=c(V1, V2), legloc=1,
xlab=xlab, ylab="Estimated RR^c(t,v)")

```

Description

Outputs to the screen a summary of the statistical tests described in Gilbert, McKeague, and Sun (2008).

Usage

```
summaryCmprsk(x)
```

Arguments

x output list from `cmprsk` function

Details

This function accepts the output list from the `cmprsk` function and returns to the screen the recommended tests and corresponding test statistics from Gilbert, McKeague and Sun (2008).

Author(s)

Peter Gilbert, Ian McKeague, and Yanqing Sun
Maintainer: Valerie Obenchain <vobencha@fhcrc.org>

References

Gilbert, P., McKeague, I. and Sun, Y. (2008) *The 2-sample problem for failure rates depending on a continuous mark: an application to vaccine efficacy*. *Biostatistics* 9, 2, 263-276.
Andersen, P. K., Borgan, O., Gill, R. D., and Keiding. (1993) *Statistical Models Based on Counting Processes*. New York: Springer.

See Also

`cmprsk` and `plotCmprsk` functions

Examples

```
## Simulate data
set.seed(100)
gp <- rep(0:1, times=50)
ftime <- 50*runif(100)
ftype <- rep(0:1, each=50)
# uninfected subjects do not have mark value; use 99
infected <- which(ftype==1)
mark <- rep(99, length(gp))
# create mark values for the infected
mark[infected] <- runif(length(infected))
# create some infected subjects with missing mark; use 99
missing <- sample(infected, 2)
mark[missing] <- 99

## Compute risk
x <- cmprsk(gp, ftime, ftype, mark, nboot=500, T1=2, T2=36, UT1=1,
```

```
UT2=42,ttanal=36)
```

```
## Summary information  
summaryCmprsk(x)
```

Index

*Topic **package**

cmprskContin-package, 2

*Topic **survival**

cmprsk, 3

plotCmprsk, 8

summaryCmprsk, 10

cmprsk, 3

cmprskContin

(*cmprskContin-package*), 2

cmprskContin-package, 2

plotCmprsk, 8

summaryCmprsk, 10