

Prevalence-Incidence Mixture Models Predictions

Description

This function produces cumulative risk predictions given an object of class PIMix a vector of times, and a data set of covariates from which to make predictions. In place of a PIMix object, all elements of the fitted model can be specified instead.

Usage

```
PIMixture.predict(x = NULL, data, time.points,
  model = "semi-parametric", prev.coef = NULL, incid.coef = NULL,
  Lambda.data = NULL, knots = NULL, order = NULL,
  spline.para.est = NULL, cov.mat = NULL, ...)
```

Arguments

<code>x</code>	object of class PIMix.
<code>data</code>	data set of covariates from which to make predictions.
<code>time.points</code>	numeric vector of times points to produce cumulative risk estimates for.
<code>model</code>	Character string indicating the specific member of the Prevalence-Incidence Mixture Model family to be fitted. Options are: <ul style="list-style-type: none"> "semi-parametric" Fits logistic regression and proportional hazards model as the prevalence and incidence models, respectively. The baseline hazard function is estimated using the iterative convex minorant algorithm. Variance estimates are obtained using bootstrap methods. Can be computationally expensive; "weakly-parametric" Fits logistic regression and proportional hazards model as the prevalence and incidence models, respectively. The baseline hazard function is approximated using integrated B-splines. "logistic-Weibull" Fits logistic regression and proportional hazards model as the prevalence and incidence models, respectively. The baseline hazard function is approximated using a Weibull distribution. "logistic-exponential" Fits logistic regression and proportional hazards model as the prevalence and incidence models, respectively. The baseline hazard function is approximated using a exponential distribution. "logistic-lognormal" Fits logistic regression and lognormal survival as the prevalence and incidence models, respectively. "logistic-loglogistic" Fits logistic regression and loglogistic survival as the prevalence and incidence models, respectively. "logistic-gengamma" Fits logistic regression and generalized-gamma survival as the prevalence and incidence models, respectively. "logistic-gamma" Fits logistic regression and gamma survival as the prevalence and incidence models, respectively.
<code>prev.coef</code>	A vector containing coefficients for the prevalence model.
<code>incid.coef</code>	A vector containing coefficients for the incidence model.
<code>Lambda.data</code>	For semi-parametric or weakly-parametric models, this is a data frame containing the times and baseline cumulative hazard.
<code>knots</code>	For weakly-parametric models, this is a numeric vector of starting time points for each exponential spline.
<code>spline.para.est</code>	For weakly-parametric models, this is a numeric vector of coefficients for each exponential spline.
<code>cov.mat</code>	A matrix containing the covariance matrix for the parameters (not required for semi-parametric models).

Value

A data frame containing the following columns

- `prev.predictor` The design matrix for the prevalence model.
- `incid.predictor` The design matrix for the incidence model.
- `time` Time.
- `CR` Cumulative risk at the time specified.
- `CR.se` Standard error for the cumulative risk.
- `LL95` Lower 95 percent confidence limit for the cumulative risk.

- LL95 Upper 95 percent confidence limit for the cumulative risk.

Author(s)

Li C. Cheung, li.cheung@nih.gov, Noorie Hyun nhyun@mcw.edu Xiaojin Xiong, Qing Pan, Hormuzd A. Katki

References

- Cheung LC, Qing P, Hyun N, Schiffman M, Fetterman B, Castle P, Lorey T, Katki H. Mixture models for undiagnosed prevalent disease and interval-censored incident disease: Applications to a cohort assembled from electronic health records. *Statistics in Medicine* 2017; 36(22):3583-95.
- Hyun N, Cheung LC, Pan Q, Katki H. Flexible risk prediction models for left or interval-censored data from electronic health records. *Annals of Applied Statistics* 11(2), 1063-1084.

Examples

```
#PIMixture includes "PIdata" RData file, and PIdata includes the two datasets, PIdata1 and PIdata2
data(PIdata)
model<- "C_CIN3PLUS+L_CIN3PLUS+R_CIN3PLUS-RES_HPV16"
fit1<-PIMixture(p.model=model,data=PIdata1, model="logistic-Weibull")
fit2<-PIMixture(p.model=model,data=PIdata1, model="weakly-parametric",n.knots=5,order=4)
fit3<-PIMixture(p.model=model,data=PIdata1, model="semi-parametric")

model2<- "C_CIN3PLUS+L_CIN3PLUS+R_CIN3PLUS-1"
fit4<-PIMixture(p.model=model2,data=PIdata1, model="non-parametric", conf.int=TRUE)
fit5<-PIMixture(p.model=model2,data=PIdata1, model="semi-parametric", conf.int=TRUE)

test.data<- data.frame(rbind(1,0))
names(test.data)<- "RES_HPV16"
time.points=c(0,12,36,60)
predict1<-PIMixture.predict(x=fit1, data=test.data, time.points=time.points)
predict2<-PIMixture.predict(x=fit2, data=test.data, time.points=time.points)
predict3<-PIMixture.predict(x=fit3, data=test.data, time.points=time.points)
predict4<-PIMixture.predict(x=fit4, data=test.data, time.points=time.points)
predict5<-PIMixture.predict(x=fit5, data=test.data, time.points=time.points)
predict1
predict2
predict3

predict4
predict5

#For stratified random samples
model3<- "C+L+R-X1+X2"
output1<-PIMixture(p.model=model3,data=PIdata2, model="semi-parametric",sample.design=1)
output2<-PIMixture(p.model=model3,data=PIdata2, model="weakly-parametric", n.knots=7,order=4,sample.design=1)
test.data<- data.frame(X1=1,X2=0.5)
time.points<-seq(0,10,by=2)
predict6<-PIMixture.predict(x=output1, data=test.data, time.points=time.points)
predict7<-PIMixture.predict(x=output2, data=test.data, time.points=time.points)
predict6
predict7
```