Package 'landmix'

October 13, 2022

Type Package

Title Landmark Prediction for Mixture Data

Version 1.0

Description Non-parametric prediction of survival outcomes for mixture data that incorporates covariates and a landmark time. Details are described in Garcia (2021) <doi:10.1093/biostatistics/kxz052>.

License GPL

Imports stats

NeedsCompilation no

Author Tanya Garcia [aut],
Layla Parast [cre]

Maintainer Layla Parast <parast@austin.utexas.edu>

Repository CRAN

Date/Publication 2022-10-10 06:00:02 UTC

R topics documented:

GenerateData *Generate data*

Description

Produces data from different populations with the probability of belonging to a population. Also produces one discrete covariate and one continuous covariate.

Usage

```
GenerateData(n, p, m, qvs, censoring.rate, simu.setting,
  covariate.dependent)
```

Arguments

`n` sample size, must be at least 1.

`p` number of populations, must be at least 2.

`m` number of different mixture proportions, must be at least 2.

`qvs` a numeric matrix of size `p` by `m` containing all possible mixture proportions (i.e., the probability of belonging to each population k, k=1,...,p.).

`censoring.rate` a scalar indicating the censoring proportion. Options are 0 or 50.

`simu.setting` Character indicating simulation setting. Options are "1A", "1B", "2A","2B". Setting "1A" and "1B" refer to Simulation setting 1 in the referenced paper, "1A" means the survival outcomes do NOT depend on the covariates, and "1B" means the survival outcomes do depend on the covariates. Setting "2A" and "2B" refer to Simulation setting 2 in the referenced paper, "2A" means the survival outcomes do NOT depend on the covariates, and "2B" means the survival outcomes do depend on the covariates.

`covariate.dependent`
logical indicator. If TRUE, then the survival times depend on covariates.

Value

Returns a list containing

- x: a numeric vector of length `n` containing the observed event times for each person in the sample.
- delta: a numeric vector of length `n` that denotes censoring (1 denotes event is observed, 0 denotes event is censored).
- q: a numeric matrix of size `p` by `n` containing the mixture proportions for each person in the sample.
- ww: a numeric vector of length `n` containing the values of the continuous covariate for each person in the sample.
- zz: a numeric vector of length `n` containing the values of the discrete covariate for each person in the sample.
- true.groups: numeric vector of length `n` denoting the population identifier for each person in the sample.

landmix.estimator — *Dynamic landmark prediction estimator for mixture data with covariates*

Description

Estimates the distribution function for mixture data where the population identifiers are unknown, but the probability of belonging to a population is known. The distribution functions are evaluated at time points `tval` and adjust for dynamic landmark prediction and one discrete covariate (`zz`) and one continuous covariate (`ww`).

Usage

```
landmix.estimator(n, m, p, qvs, q, x, delta, ww, zz, run.NPNA,
  run.NPNA_avg, tval, tval0, z.use, w.use)
```

Arguments

`n` sample size, must be at least 1.

`m` number of different mixture proportions, must be at least 2.

`p` number of populations, must be at least 2.

`qvs` a numeric matrix of size `p` by `m` containing all possible mixture proportions (i.e., the probability of belonging to each population k, k=1,...,p.).

`q` a numeric matrix of size `p` by `n` containing the mixture proportions for each person in the sample.

`x` a numeric vector of length `n` containing the observed event times for each person in the sample.

`delta` a numeric vector of length `n` that denotes censoring (1 denotes event is observed, 0 denotes event is censored).

`ww` a numeric vector of length `n` containing the values of the continuous covariate for each person in the sample.

`zz` a numeric vector of length `n` containing the values of the discrete covariate for each person in the sample.

`run.NPNA` a logical indicator. If TRUE, then the output includes the estimated distribution function for mixture data that accounts for covariates and dynamic landmarking. This estimator is called "NPNA" in the referenced paper.

`run.NPNA_avg` a logical indicator. If TRUE, then the output includes the estimated distribution function for mixture data that averages out over the observed covariates. This is referred to as NPNA_marg in the referenced paper.

`tval` numeric vector of time points at which the distribution function is evaluated, all values must be non-negative.

`tval0` numeric vector of time points representing the landmark times. All values must be non-negative and smaller than the maximum of `tval`.

`z.use` numeric vector at which to evaluate the discrete covariate Z at in the estimated distribution function. The values of `z.use` must be in the range of the observed `zz`.

`w.use` numeric vector at which to evaluate the continuous covariate W at in the estimated distribution function. The values of `w.use` must be in the range of the observed `ww`.

Value

`landmix.estimator` returns a list containing

- Ft.estimate: a numeric array containing the estimated distribution functions for all methods for all `p` populations. The distribution function is evaluated at each `tval`, `tval0`, `z.use`, `w.use`, and for all `p` populations. The dimension of the array is the number of methods by `length(tval)` by `lenth(tval0)` by `length(z.use)` by `length(w.use)` by `p`. The distribution function is only valid for $t \geq t_0$, so `Ft.estimate` shows NA for any combination for which $t < t_0$.
- St.estimate: a numeric array containing the estimated distribution functions for all methods for all `m` mixture proportion subgroups. The distribution function is evaluated at each `tval`, `tval0`, `z.use`, `w.use`, and for all `m` mixture proportion subgroups. The dimension of the array is the number of methods by `length(tval)` by `lenth(tval0)` by `length(z.use)` by `length(w.use)` by `m`. The distribution function is only valid for $t \geq t_0$, so `St.estimate` shows NA for any combination for which $t < t_0$.

Details

We estimate the distribution function for mixture data where the population identifiers are unknown, but the probability of belonging to a population is known. The distribution functions are evaluated at time points `tval` and adjust for dynamic landmark prediction and one discrete covariate (`zz`) and one continuous covariate (`ww`). Dynamic landmark prediction means that the distribution function is computed knowing that the survival time, T, satisfies $T > t_0$ where t_0 are the time points in `tval0`.

Examples

```
# Setup parameters to generate the data
set.seed(1)
censoring.rate <- 40
p <- 2
n <- 2000
m <- 4
tval <- seq(0,80,by=5)
tval0 <- c(0,20,30,40,50)
z.use <- c(0,1)
w.use <- seq(35,55,by=1)
simu.setting <- "2A"
covariate.dependent <- TRUE
run.NPMLEs <- TRUE
run.NPNA <- TRUE
run.OLS <- FALSE
run.WLS <- FALSE
run.EFF <- FALSE
run.NPNA_avg <- FALSE


## compute the finite set of mixture proportions
qvs <- qvs.values(p,m)

## generate the data

data.gen <- GenerateData(n,p,m,qvs,censoring.rate,simu.setting,covariate.dependent)

x <- data.gen$x
delta <- data.gen$delta
q <- data.gen$q
ww <- data.gen$ww
zz <- data.gen$zz


## true group membership (needed to compute the AUC/BS for simulated data
true.groups <- data.gen$true.groups

## Perform the estimation
estimators.out <- landmix.estimator(n,m,p,qvs,q,
x,delta,ww,zz,
run.NPNA,
run.NPNA_avg,
tval,tval0,
z.use,w.use)
```

`qvs.values` *Generate finite set of mixture proportions*

Description

Produces the finite set of mixture proportions for simulated data.

Usage

```
qvs.values(p, m)
```

Arguments

`p`	number of populations, must be at least 2.
`m`	number of different mixture proportions, must be at least 2.

Value

Returns a `p` by `m` matrix of mixture proportions.

Index