Package 'Icens'

February 21, 2024

Title NPMLE for Censored and Truncated Data

Description Many functions for computing the NPMLE for censored and truncated data.

Version 1.75.0

Author R. Gentleman and Alain Vandal

Maintainer Bioconductor Package Maintainer

<maintainer@bioconductor.org>

Depends survival

Imports graphics

License Artistic-2.0

biocViews Infrastructure

RoxygenNote 7.2.3

git_url https://git.bioconductor.org/packages/Icens

git_branch devel

git_last_commit 233ce54

git_last_commit_date 2023-10-24

Repository Bioconductor 3.19

Date/Publication 2024-02-21

Contents

`Bisect` *An implementation of the bisection algorithm for root finding.*

Description

Most of the optimizations in *Icens* have a one dimensional root-finding component. Since the quantities involved are generally restricted to a subset of [0,1] we use bisection to find the roots.

Usage

```
Bisect(tA, pvec, ndir, Meps, tolbis=1e-07)
```

Arguments

`tA`	The transpose of the clique matrix.
`pvec`	The current estimate of the probability vector.
`ndir`	The direction to explore.
`Meps`	Machine epsilon, elements of `pvec` that are less than this are assumed to be zero.
`tolbis`	The tolerance used to determine if the algorithm has converged.

Details

We search from `pvec` in the direction `ndir` to obtain the new value of `pvec` that maximizes the likelihood.

Value

The new estimate of `pvec`.

Author(s)

Alain Vandal and Robert Gentleman.

References

Any book on optimization.

`BVcliques` *Find the bivariate cliques from the marginal data.*

Description

The maximal cliques of the intersection graph are obtained by first finding the cliques for the marginal data and then combining them using the algorithm in Gentleman and Vandal (1999).

Usage

```
BVcliques(intvlx, intvly, Lxopen=TRUE, Rxopen=FALSE,
                    Lyopen=TRUE, Ryopen=FALSE )
```

Arguments

`intvlx`	The cliques for one marginal component, alternatively the marginal intervals can be supplied.
`intvly`	The cliques for the other marginal component, alternatively the marginal intervals can be supplied.
`Lxopen`	Boolean indicating whether the left end point in the x coordinate is open.
`Rxopen`	Boolean indicating whether the right end point in the x coordinate is open.
`Lyopen`	Boolean indicating whether the left end point in the y coordinate is open.
`Ryopen`	Boolean indicating whether the right end point in the y coordinate is open.

Value

A list of the maximal cliques of the intersection graph of the data.

Author(s)

A. Vandal and R. Gentleman

References

Graph–Theoretical Aspects of Bivariate Censored Data, R. Gentleman and A. Vandal, 1999, submitted.

See Also

`BVclmat`, `BVsupport`

Examples

```
data(cmv)
cmv.cl <- BVcliques(cmv[,1:2], cmv[,3:4], Lxopen=FALSE, Lyopen=FALSE )
```

BVclmat *Compute the clique matrix from the clique list.*

Description

Given the clique list, obtained from BVcliques, the clique matrix is obtained. This is the m (number of cliques) by n (number of observations) matrix. A[i,j] is one if individual j is in maximal clique i.

Usage

```
BVclmat(cliques)
```

Arguments

cliques	The clique list.

Value

The m by n clique matrix.

Author(s)

A. Vandal and R. Gentleman

References

Graph–Theoretical Aspects of Bivariate Censored Data, R. Gentleman and A. Vandal, 1999, submitted.

See Also

BVcliques, BVsupport

Examples

```
data(cmv)
bcl <- BVcliques(cmv[,1:2], cmv[,3:4])
A <- BVclmat(bcl)
```

BVsupport *Compute the support for the cliques of a bivariate intersection graph.*

Description

Given the regions where the events occurred and the cliques of the intersection graph the support of the cliques is computed. For each clique it is the intersection of the event time regions for all observations in that clique.

Usage

```
BVsupport(intvlx, intvly, cliques=BVcliques(intvlx, intvly))
```

Arguments

intvlx	The event time intervals for one dimension.
intvly	The event time intervals for the other dimension.
cliques	The list of maximal cliques of the intersection graph, optionally.

Value

An m by 4 matrix containing the corners of the intervals of support for the maximal cliques of the intersection graph corresponding to the first two arguments to the function.

Author(s)

A. Vandal and R. Gentleman

References

Graph–Theoretical Aspects of Bivariate Censored Data, R. Gentleman and A. Vandal, 1999, submitted.

See Also

BVcliques, BVclmat

Examples

```
data(cmv)
cmv.cl <- BVcliques(cmv[,1:2], cmv[,3:4])
boxes <- BVsupport(cmv[,1:2], cmv[,3:4], cmv.cl)
```

cmv — *Data on times to shedding of cytomegalovirus and to colonization of mycobacterium avium complex.*

Description

The cmv data frame has 204 rows and 4 columns. The intervals should be treated as closed at both ends to replicate the analysis in Betensky and Finkelstein.

Format

This data frame contains the following columns:

cmvL The left end of the CMV shedding interval.

cmvR The right end of the CMV shedding interval.

macL The left end of the MAC colonization interval.

macR The right end of the MAC colonization interval.

Details

Betensky and Finkelstein, 1999 present data from the AIDS Clinical Trials Group protocol ACTG 181. This was a natural history substudy of a comparative trial. Patients were scheduled for clinic visits during follow–up and data was collected on the time until two events; shedding of cytomegalovirus (CMV) in the urine and blood and for colonization of mycobacterium avium complex (MAC) in the sputum or stool.

Source

Betensky, R. A. and Finkelstein, D. M., 1999, *A nonparametric maximum likelihood estimator for bivariate interval censored data*, Statistics in Medicine,

Examples

```
data(cmv)
```

cosmesis — *The time taken until cosmetic deterioration of breast cosmesis.*

Description

The cosmesis data frame has 95 rows and 3 columns.

Format

This data frame contains the following columns:

L The left end point of the cosmetic deterioration interval.

R The right end point of the cosmetic deterioration interval.

Trt The treatment indicator. It is zero for those that received radiotherapy.

Source

A semiparametric model for regression analysis of interval-censored failure time data, D. M. Finkelstein and R. A. Wolfe, 1985, Biometrics.

Examples

```
data(cosmesis)
```

`EM` *A function to compute the NPMLE of p based on the incidence matrix A.*

Description

The incidence matrix, `A` is the m by n matrix that represents the data. There are m probabilities that must be estimated. The EM, or expectation maximization, method is applied to these data.

Usage

```
EM(A, pvec, maxiter=500, tol=1e-12)
```

Arguments

`A`	The incidence matrix.
`pvec`	The probability vector.
`maxiter`	The maximum number of iterations.
`tol`	The tolerance used to judge convergence.

Details

Lots.

Value

An object of class `icsurv` containing the following components:

`pf`	The NPMLE of the probability vector.
`numiter`	The number of iterations used.
`converge`	A boolean indicating whether the algorithm converged.
`intmap`	If present indicates the real representation of the support for the values in `pf`.

Author(s)

Alain Vandal and Robert Gentleman.

References

The EM algorithm applied to the maximal cliques of the intersection graph of the censored data. *The empirical distribution function with arbitrarily grouped, censored and truncated data*, B. W. Turnbull, 1976, JRSS;B.

See Also

VEM, ISDM, EMICM, PGM

Examples

```
data(cosmesis)
csub1 <- subset(cosmesis, subset= Trt==0, select=c(L,R))
EM(csub1)
data(pruitt)
EM(pruitt)
```

EMICM *Compute the NPMLE for censored data using the EMICM.*

Description

An implementation of the hybrid EM ICM (Iterative convex minorant) estimator of the distribution function proposed by Wellner and Zahn (1997).

Usage

```
EMICM(A, EMstep=TRUE, ICMstep=TRUE, keepiter=FALSE, tol=1e-07,
maxiter=1000)
```

Arguments

A	Either the m by n clique matrix or the n by 2 matrix containing the event time intervals.
EMstep	Boolean, indicating whether to take an EM step in the iteration.
ICMstep	Boolean, indicating whether to take an ICM step.
keepiter	Boolean determining whether to keep the iteration states.
tol	The maximal L1 distance between successive estimates before stopping iteration.
maxiter	The maximal number of iterations to perform before stopping.

Details

Lots, and they're complicated too!

Value

An object of class `icsurv` containing the following components:

`pf`	The estimated probabilities.
`sigma`	The NPMLE of the survival function on the maximal antichains.
`weights`	The diagonal of the likelihood function's second derivative.
`lastchange`	A vector of differences between the last two iterations.
`numiter`	The total number of iterations performed.
`iter`	Is only present if `keepiter` is true; states of sigma during the iteration.
`intmap`	The real representation associated with the probabilities reported in `pf`.

Author(s)

Alain Vandal and Robert Gentleman

References

A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data, J. A. Wellner and Y. Zhan, 1997, JASA.

See Also

`EM`,`VEM`, `PGM`

Examples

```
data(cosmesis)
csub1 <- subset(cosmesis, subset=Trt==0, select=c(L,R))
EMICM(csub1)
data(pruitt)
EMICM(pruitt)
```

---

hiv *Intervals for infection time and disease onset for 257 hemophiliac patients.*

Description

The `hiv` data frame has 257 rows and 4 columns.

Format

This data frame contains the following columns:

yL The left end point of the infection time interval.

yR The right end point of the infection time interval.

zL The left end point of the disease onset interval.

zR The right end point of the disease onset interval.

Age Coded as 1 if the estimated age at infection was less than 20 and 2 if the estimated age at infection was greater than 20.

Trt Treatment, Light or Heavy

Details

The setting is as follows. Individuals were infected with the HIV virus at some unknown time they subsequently develop AIDS at a second unknown time. The data consist of two intervals, (y_L, y_R) and (z_L, z_R), such that the infection time was in the first interval and the time of disease onset was in the second interval. A quantity of interest is the incubation time of the disease which is $T = Z - Y$. The authors argue persuasively that this should be considered as bivariate interval censored data. They note that simply forming the differences $(z_L - y_R, z_R - y_L)$ and analysing the resultant data assumes an incorrect likelihood. DeGruttola and Lagakos transform the problem slightly to study the joint distribution of Y and $T = Z - Y$. This is equivalent to estimating the joint distribution of Z and Y then transforming. The data, as reported, have been discretized into six month intervals.

We use the data as reported in Table 1 of DeGruttola and Lagakos, 1989. The patients were 257 persons with Type A or B hemophilia treated at two hospitals in France. They were then examined intermittently (as they came in for treatment?) and their HIV and AIDS status was determined. Kim, De Gruttola and Lagakos report some covariate information and their paper is concerned with the modeling of that information. In this paper we concentrate only on the event times and ignore the covariate information; that topic being worthy of separate investigation.

Source

DeGruttola, V. and Lagakos, S.W., 1989, *Analysis of doubly-censored survival data, with application to AIDS*, Biometrics.

Kim, Mimi Y. and De Gruttola, Victor G. and Lagakos, Stephen W., 1993, *Analyzing Doubly Censored Data With Covariates, With Application to AIDS*, Biometrics.

Examples

```
data(hiv)
```

`Icens-internal` *Internal Icens functions*

Description

Internal Icens functions

Details

These are not to be called by the user.

`icsurv` *The class of objects returned by the estimation routines in the Icens library.*

Description

An object of class `icsurv` must contain the following components:

converge A boolean indicating whether the iteration producing `pf` converged.

pf The probability vector.

It can optionally contain any of the following components:

clmat The clique matrix used to obtain `pf`.

intmap The real representations of the support for the components of `pf`.

iter A matrix containing every iterative estimate of `pf`, useful for debugging.

lval The value of the **log** likelihood at `pf`.

numiter The number of iterations taken.

sigma The cumulative sum of `pf`.

weights Weights used in the EMICM algorithm.

Author(s)

Alain Vandal and Robert Gentleman.

See Also

`VEM`, `ISDM`, `EMICM`, `PGM`, `EM`

ISDM *Estimate the NPMLE of censored data using the ISDM method proposed in Lesperance and Kalbfleisch (19*

Description

ISDM is a method for estimating the NPMLE of censored data.

Usage

```
ISDM(A, pvec, maxiter=500, tol=1e-07, tolbis=1e-08, verbose=FALSE)
```

Arguments

A	The m by n incidence, or clique, matrix. Or the n by 2 matrix containing the event intervals.
pvec	An initial estimate of the probability vector; not required.
maxiter	Maximum number of iterations to be made.
tol	The tolerance used to determine convergence.
tolbis	A second tolerance used for the steps.
verbose	Boolean, should verbose output be printed.

Details

Lots of complicated stuff should go here.

Value

A list containing:

pf	The estimated NPMLE of the probability vector.
numiter	The number of iterations performed.

Author(s)

Alain Vandal and Robert Gentleman

References

An Algorithm for Computing the Nonparametric MLE of a Mixing Distribution, Lesperance, Mary L. and Kalbfleisch, John D., JASA, 1992

See Also

VEM, EMICM, PGM

Examples

```
    data(cosmesis)
    csub1 <- subset(cosmesis, subset=Trt==0, select=c(L,R))
    ISDM(csub1)
#    data(pruitt)
#    ISDM(pruitt)
```

`Maclist` *A function to*

Description

Returns a list of maximal cliques of the intersection graph of the real valued intervals supplied in `m`. These are one dimensional intervals with one interval for each individual. The algorithm is coded in interpreted code and should be moved to compiled code for speed. How do we handle exact failure times? Which algorithm is used?

Usage

```
Maclist(intvls, Lopen=TRUE, Ropen=FALSE)
```

Arguments

`intvls`	A n by 2 matrix, the first column is the left endpoints and the second column contains the right endpoints of the failure time intervals.
`Lopen`	A boolean indicating whether the intervals are open on the left.
`Ropen`	A boolean indicating whether the intervals are open on the right.

Value

A list of length m. Each element of the list corresponds to one maximal antichain. The row numbers (from `m`) identify the individuals and all row numbers for the individuals in the maximal clique. Maximal cliques occur in their natural (left to right) order.

Author(s)

Alain Vandal and Robert Gentleman

References

Computational Methods for Censored Data using Intersection Graphs, R. Gentleman and A. Vandal, JCGS, 2000.

See Also

`Macmat`

Examples

```
data(cosmesis)
csub1 <- subset(cosmesis, subset=Trt==0, select=c(L,R))
ml1 <- Maclist(csub1)
```

Macmat *A function to compute the incidence matrix for an intersection graph.*

Description

Returns the Petrie matrix and Petrie pairs of an interval order given its list of maximal antichains. These can be obtained from Maclist.

Usage

```
Macmat(ml)
```

Arguments

ml	A list containing the maximal cliques of the intersection graph of the data.

Details

Not worth mentioning?

Value

A list containing two components.

pmat	The Petrie or clique matrix of the underlying interval order.
ppairs	The Petrie pairs for each observation. These indicate the first and last maximal clique occupied by the observation.

Author(s)

Alain Vandal and Robert Gentleman

References

Computational Methods for Censored Data using Intersection Graphs, R. Gentleman and A. Vandal, JCGS, 2000.

See Also

Maclist

Examples

```
data(cosmesis)
csub1 <- subset(cosmesis, subset=Trt==0, select=c(L,R))
ml1 <- Maclist(csub1)
mm1 <- Macmat(ml1)
```

MLEintvl *Compute the real representation for the maximal cliques.*

Description

The intervals on the real line that corresponds to the intersections of the maximal cliques are computed and returned.

Usage

```
MLEintvl(intvls, ml=Maclist(intvls))
```

Arguments

intvls	The n by 2 matrix containing the event time intervals for the individuals under study.
ml	The Maclist computed for the intvls.

Value

An m by 2 matrix, where m is the number of maximal cliques. The first column contains the left end point of the real representation for the appropriate maximal clique and the second column contains the right end point.

Author(s)

Alain Vandal and Robert Gentleman

References

Computational Methods for Censored Data using Intersection Graphs, R. Gentleman and A. Vandal, JCGS, 2000.

See Also

Maclist

Examples

```
data(cosmesis)
csub1 <- subset(cosmesis, subset=Trt==0, select=c(L,R))
MLEintvl(csub1)
```

PGM *An implementation of the projected gradient methods for finding the NPMLE.*

Description

An estimate of the NPMLE is obtained by using projected gradient methods. This method is a special case of the methods described in Wu (1978).

Usage

```
PGM(A, pvec, maxiter = 500, tol=1e-07, told=2e-05, tolbis=1e-08,
    keepiter=FALSE)
```

Arguments

A	A is either the m by n clique matrix or the n by 2 matrix containing the left and right end points for each event time.
pvec	An initial estimate of the probability vector.
maxiter	The maximum number of iterations to take.
tol	The tolerance for decreases in likelihood.
told	told does not seem to be used.
tolbis	The tolerance used in the bisection code.
keepiter	A boolean indicating whether to return the number of iterations.

Details

New directions are selected by the projected gradient method. The new optimal pvec is obtained using the bisection algorithm, moving in the selected direction. Convergence requires both the L_1 distance for the improved pvec and the change in likelihood to be below tol.

Value

An object of class icsurv containing the following components:

pf	The NPMLE of pvec.
sigma	The cumulative sum of pvec.
lval	The value of the log likelihood at pvec.
clmat	The clique matrix.
method	The method used, currently only "MPGM" is possible.
lastchange	The difference between pf and the previous iterate.
numiter	The number of iterations carried out.
eps	The tolerances used.
converge	A boolean indicating whether convergence occurred within maxiter iterations.
iter	If keepiter is true then this is a matrix containing all iterations - useful for debugging.

Author(s)

Alain Vandal and Robert Gentleman.

References

Some Algorithmic Aspects of the Theory of Optimal Designs, C.–F. Wu, 1978, Annals.

See Also

VEM, ISDM, EMICM, PGM, EM

Examples

```
data(cosmesis)
csub1 <- subset(cosmesis, subset=Trt==0, select=c(L,R))
PGM(csub1)
data(pruitt)
PGM(pruitt)
```

---

plot.icsurv *A plot method for the estimates produced by the estimation methods in Icens.*

Description

Produces nice plots of the estimated NPMLE.

Usage

```
## S3 method for class 'icsurv'
plot(x, type="eq", surv=FALSE, bounds=FALSE, shade=3, density=30,
angle=45, lty=1, new=TRUE, xlab="Time", ylab="Probability", main="GMLE",
ltybnds=2, ...)
```

Arguments

x	The estimate of the NPMLE.
type	Three options, "eq" for equivalence call, "gw" for the Groeneboom-Wellner estimate, and "lc" for the left-continuous estimate.
surv	Logical indicating whether or not to plot the survival curve.
bounds	Logical indicating whether or not to include bounds around the estimate.
shade	An integer in 1, 2, or 3 denoting what component of the plot to shade.
density	The density of shading lines, in lines per inch.
angle	The slope of shading lines, given as an angle in degrees (counter-clockwise).
lty	The line type for the estimates.

new	Logical indicating whether or not to create a new plot.
xlab	The x-axis label.
ylab	The y-axis label.
main	The main title for the plot.
ltybnds	The line type for the bounds on the estimates.
...	Additional arguments passed to the plot function.

Value

No value is returned. A plot of the NPMLE is made on the active graphics device.

Author(s)

Alain Vandal and Robert Gentleman.

See Also

VEM, ISDM, EMICM, PGM

Examples

```
data(cosmesis)
csub1 <- subset(cosmesis, subset=Trt==0, select=c(L,R))
e1 <- VEM(csub1)
par(mfrow=c(2,2))
plot(e1)
data(pruitt)
e2 <- EM(csub1)
plot(e2)
e3 <- PGM(csub1)
plot(e3)
e4 <- EMICM(csub1)
plot(e4)
```

Plotboxes *Plot the event time regions for bivariate data.*

Description

Plot rectangles described by the interval given in the first two arguments.

Usage

```
Plotboxes(int1, int2, textp=FALSE, showmac=FALSE, showsupp=FALSE, showmp=FALSE,
cliques=NULL, macprod=NULL, density=c(2, 8, 20), col=c(2, 3, 4),
offsetx=0.02, offsety=0.03)
```

Arguments

int1	The intervals for the x dimension.
int2	The intervals for the y dimension.
textp	Boolen, if true add text.
showmac	Boolean, if true then the maximal cliques are shown in a different color?
showsupp	Boolean, if true show support boxes.
showmp	Boolean
cliques	Maximal cliques.
macprod	macprod
density	The density of the polygon shading lines, in lines per inch.
col	Color for plotting features.
offsetx	Offset for x-axis.
offsety	Offset for y-axis.

Value

No value is returned. The event rectangles are plotted on the active graphics device.

Author(s)

A. Vandal and R. Gentleman

References

Graph–Theoretical Aspects of Bivariate Censored Data, R. Gentleman and A. Vandal, 1999, submitted.

See Also

BVclmat, BVsupport, BVcliques

Examples

```
data(cmv)
Plotboxes(cmv[,1:2], cmv[,3:4], showmac=TRUE)
```

PMGA *Implement the pool monotone groups algorithm.*

Description

For isotonization problems some increase in speed and decrease in complexity can be achieved through the use of the pool monotone groups algorithm of Y.L. Zhang and M.A. Newton (1997). It isotonizes a weighted and ordered set of values.

Usage

```
PMGA(est, ww=rep(1, length(est)))
```

Arguments

`est`	The vector of values, in the appropriate order.
`ww`	The weight vector.

Details

To be supplied at some later date.

Value

An object containing the following components:

`est`	The isotonized estimates.
`ww`	The weights associated with the isotonized estimates.
`poolnum`	The number of values pooled in the current estimate.
`passes`	The number of passes which were required to isotonize the list.

Author(s)

Alain Vandal and Robert Gentleman.

References

Y.L. Zhang and M.A. Newton (1997), http://www.stat.wisc.edu/~newton/newton.html)

See Also

EMICM

pruitt *A small artificial, bivariate right-censored data set.*

Description

The pruitt data was given in Pruitt (1993) as an example for testing different methods of estimating the bivariate NPMLE for right censored data. This matrix represents the clique matrix of the intersection graph of the data set given by Pruitt.

Format

This data frame contains 8 columns, labeled A through H that represent the observations. There are seven rows corresponding to the seven maximal cliques in the intersection graph.

Source

Small Sample Comparison of Six Bivariate Survival Curve Estimators, Journal of Statistical Computation and Simulation, R. Pruitt, 1993.

Examples

```
data(pruitt)
```

VEM *Compute the NPMLE of p via the Vertex Exchange Method.*

Description

The Vertex Exchange Method is used to obtain the NPMLE of p.

Usage

```
VEM(A, pvec, maxiter=500, tol=1e-07, tolbis=1e-07, keepiter=FALSE)
```

Arguments

A	The m by n incidence matrix or the n by 2 matrix of intervals.
pvec	The initial estimate for the probability vector.
maxiter	The maximum number of iterations allowed.
tol	The tolerance used to determine convergence.
tolbis	The tolerance used in the bisection stage of the algorithm.
keepiter	Should iteration information be retained and returned.

Details

Lots.

Value

An object of class `icsurv` with the following components.

`pf`	The NPMLE of the probability vector.
`numiter`	The number of iterations used.
`lval`	The value of the logarithm of the likelihood at the NPMLE.
`converge`	Boolean stating whether the iteration converged.
`intmap`	If present it contains the real representations for the maximal cliques. These are the intervals (on the real line) where the mass in `pf` is placed.

Author(s)

Robert Gentleman and Alain Vandal

References

A Vertex-exchange-method in D-optimal Design Theory , D. Bohning, Metrika, 1986.

See Also

`EM`, `ISDM`, `EMICM`, `PGM`

Examples

```
data(cosmesis)
csub1 <- subset(cosmesis, subset=Trt==0, select=c(L,R))
VEM(csub1)
data(pruitt)
VEM(pruitt)
```

Index