Package ‘dblcens’

April 17, 2009

Title Compute the NPMLE of distribution from doubly censored data

Version 1.1.4

Author Mai Zhou, Li Lee, Kun Chen.

Description Use EM algorithm to compute the NPMLE of CDF and also the two censoring distributions. For doubly censored data (as described in Chang and Yang (1987) Ann. Stat. 1536-47). You can also specify a constraint, it will return the constrained NPMLE and the -2 log empirical likelihood ratio. This can be used to test the hypothesis about the constraint and find confidence intervals for F(T) via empirical likelihood ratio theorem. Influence function of hat F may also be calculated (but may be slow).

Maintainer Mai Zhou <mai@ms.uky.edu>

License GPL (>= 2)

Repository CRAN

Date/Publication 2008-04-04 20:39:46

R topics documented:

d011 1
d011ch 4

Index **8**

`d011` *Compute NPMLE of CDF from doubly censored data*

Description

`d011` computes the NPMLE of CDF from doubly censored data via EM algorithm starting from an initial estimator that have jumps at (1) uncensored points; (2) (mid-point of) consecutive survival times with censoring indicator pattern of (0,2), (see below for definition)

When there are ties, the left (right) censored points are treated as happened before (after), to break tie. Also when the last obs. happens to be right censored observation and when the first obs. happens to be left censored observations are changed to uncensored, in order to obtain a proper distribution as the CDF estimator. (though this can be modified easily as they are written in R language).

It also computes the NPMLE of the two censoring distributions. There is an option that you may also try to compute the three influence functions.

Usage

```
d011(z, d, identical = rep(0, length(z)),
         maxiter = 49, error = 0.00001, influence.fun = FALSE)
```

Arguments

`z`	a vector of length n denoting observed times, (ties permitted)
`d`	a vector of length n that contains censoring indicator: d= 2 or 1 or 0, (according to z being left, not, right censored)
`identical`	optional. A vector of length n that has values either 0 or 1. identical[i]=1 means even if $(z[i], d[i])$ is identical with $(z[j], d[j])$, for some $j \neq i$, they still stay as 2 observations, (not 1 obs. with weight 2, which only happen if identical[i]=0 and identical[j] =0). One reason for this is because they may have different covariates not shown here. This adds more flexibility for regression applications. Default value is identical = 0, (i.e. collapse if identical observations).
`maxiter`	optional integer value. default to 49
`error`	optional. Default to 0.00001
`influence.fun`	optional. Default to FALSE. If TRUE, the code will try to compute the influence functions (3 of them) at the censored times. This computation can be very slow and memory intensive (for data with >500 censored times).

Value

a list contain the NPMLE of CDF and other informations.

`time`	Times of input z, with time corresponding to status=2 removed.
`status`	Censoring status of the above times. Status = -1 means this is an added time because of the censoring pattern (0,2).
`surv`	Survival probability at the above times.
`jump`	Jumps of the NPMLE at the above times.
`exttime`	Similar to times but those with status =2 not removed.
`extstatus`	status of exttime

`extjump`	jump pf NPMLE at exttime.
`extsurv.Sx`	Estimated lifetime distribution.
`surv0.Sy`	One of the censoring distributions.
`jump0`	Jump of surv0.Sy
`surv2.Sz`	Another censoring distribution.
`jump2`	Jump of surv2.Sz
`conv`	A vector of length 2: the actual number of iterations, and the actual error of successive iteration. If the iteration number equal to the maxiter you set, then the iteration has not converged.
`Nodes`	Points where the influence function is computed.
`IC1tu`	Influence function value at the nodes. See Chang (1990) for details.
`IC1tu2`	Influence function value at other points. See Chang (1990) for details.
`IC2tu`	ditto IC1tu
`IC3tu`	ditto IC1tu
`VarFt`	Estimated variances of F(t) at the Nodes.

Author(s)

Mai Zhou ⟨mai@ms.uky.edu⟩, Li Lee.

References

Chang, M. N. and Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. Ann. Statist. 15, 1536-1547.

Turnbull (1976) The empirical distribution function with arbitrarily grouped, censored and truncated data. JRSS B 290-295.

Chang, M. N. (1990). Weak convergence in doubly censored data. Ann. Statist. 18, 390-405.

Chen, K. and Zhou, M. (2000). Nonparametric Hypothesis Testing and Confidence Intervals with Doubly Censored Data. Tech. Report, Univ. of Kentucky. See also Lifetime Data Analysis, **9**, (2003).

Examples

```
d011(z=c(1,2,3,4,5), d=c(1,0,2,2,1))
#
# you should get something like below (and more)
#
#       $time:
#       [1] 1.0 2.0 2.5 5.0    (notice the times, (3,4), corresponding
#                                   to d=2 are removed, and time 2.5 added
#       $status:               since there is a (0,2) pattern at
#       [1]  1  0 -1  1        times 2, 3. The status indicator of -1
#                                   show that it is an added time )
#       $surv
#       [1] 0.5000351 0.5000351 0.3333177 0.0000000
#
```

```
#       $jump
#       [1] 0.4999649 0.0000000 0.1667174 0.3333177
#
#       $exttime
#       [1] 1.0 2.0 2.5 3.0 4.0 5.0
#
#       $extstatus
#       [1]  1  0 -1  2  2  1
#
#       ......
#
#       $conv
#       [1] 3.300000e+01  8.788214e-06  ### did 33 iterations
#
# BTW, the true NPMLE of surv is (1/2, 1/2, 1/3, 0) at times (1,2,2.5,5).
###### Example 2.
d011(c(1,2,3,4,5), c(1,2,1,0,1),influence.fun=TRUE)
#     we get
# ......
#$conv:
#[1] 3 0
#
#$Nodes:
#[1] 2 4
#
#$IC1tu:
#     [,1] [,2]
#[1,]   -1    0
#[2,]   -1   -2
#
#$IC2tu:
#           [,1] [,2]
#[1,]  0.0000000    0
#[2,] -0.3333333    0
#
#$IC3tu:
#     [,1]       [,2]
#[1,]   -1 -0.6666667
#[2,]   -1 -1.0000000
#
#$VarFt:
#[1] 0.24 0.24           ## est var of F(t) at t=nodes
#######################################################
```

d011ch *Compute NPMLE of CDF from doubly censored data, with a constraint*

Description

`d011ch` computes the NPMLE of CDF, with and without a constraint, from doubly censored data.

It also computes the -2 log empirical likelihood ratio for testing the constraint via empirical likelihood theorem, i.e. under Ho it should be distributed as chi-square with df=1.

It uses EM algorithm starting from an initial CDF estimator that have jumps at uncensored points as well as the mid-point of those censoring times that have a pattern of (0,2), (see below for definition and example.)

The constraint on the CDF are given in the form F(T) = p where you specify the time T and probability p.

When there are ties among censored and uncensored observations, the left (right) censored points are treated as happened before (after), to break tie. Also the last right censored observation and first left censored observations are changed to uncensored, in order to obtain a proper distribution as estimator. (though this can be modified easily as they are written in R language).

Usage

```
d011ch(z, d, K, konst,
     identical = rep(0, length(z)), maxiter = 49, error = 0.00001)
```

Arguments

`z`	a vector of length n denoting observed times, (ties permitted)
`d`	a vector of length n that contains censoring indicator: d= 2 or 1 or 0, (according to z being left, not, right censored)
`K`	the constraint time.
`konst`	the constraint value, i.e. F(K)=konst.
`identical`	optional. a vector of length n that has values either 0 or 1. identical[i]=1 means even if $(z[i], d[i])$ is identical with $(z[j], d[j])$, for some $j \neq i$, they still stay as 2 observations, not 1 obs. with weight 2, which only happen if identical[i]=0 and identical[j] =0. One reason to do this is because they may have different covariates not shown here. This flexibility may be useful for regression applications. Default value is identical = 0.
`maxiter`	optional integer value. Default to 49
`error`	optional. Default to 0.00001

Value

a list contain the NPMLE of CDF with and without the constraint, -2loglik ratio and other informations.

`time`	survival times. Those corresponding to d=2 are removed. Those corresponding to (0,2) censoring pattern are added, at mid-point.
`status`	Censoring status of the above times. Since left censored times are removed, there is no status =2. There may be -1, indicating that this is an added time for (0,2) censoring pattern.
`surv`	The survival function at the above times.
`jump`	Jumps of NPMLE at the above times.
`exttime`	Similar to time but now include the left censored times.

`extstatus`	Censoring status of exttime. -1 has same meaning as status before.
`extjump`	Jumps of the unconstrained NPMLE on extended times.
`extsurv.Sx`	Survival probability at exttime.
`konstdist`	The constrained NPMLE of distribution.
`konstjump`	Jumps of the constrained NPMLE of CDF.
`konsttime`	Location of the constraint, same as K in the input.
`theta`	is the same value `konst` in the input.
`"-2loglikR"`	the Wilks statistics. Distributed approximately chi-square df=1 under Ho
`maxiter`	the actual number of iterations for the unconstrained NPNLE. The constrained NPMLE usually took less iterations to converge.

Author(s)

Kun Chen, Mai Zhou ⟨mai@ms.uky.edu⟩

References

Chang, M. N. and Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. Ann. Statist. 15, 1536-1547.

Murphy, S. and Van der Varrt. (1997). Semiparametric Likelihood Ratio Inference. Ann. Statist. 25, 1471-1509.

Chen, K. and Zhou, M. (2000). Nonparametric Hypothesis Testing and Confidence Intervals with Doubly Censored Data. Tech Report, Univ. of Kentucky. This paper appeared in: Lifetime Data Analysis (2003). **9**

Examples

```
d011ch(z=c(1,2,3,4,5), d=c(1,0,2,2,1), K=3.5, konst=0.6)
#
# Here we are testing Ho: F(3.5) = 0.6 with a two-sided alternative
# you should get something like
#
#       $time:
#       [1] 1.0 2.0 2.5 5.0    (notice the times, (3,4), corresponding
#                                   to d=2 are removed, and time 2.5 added
#       $status:               since there is a (0,2) pattern at
#       [1]  1  0 -1  1        times 2, 3. The status indicator of -1
#                                   show that it is an added time )
#       $surv
#       [1] 0.5000351 0.5000351 0.3333177 0.0000000
#
#       $jump
#       [1] 0.4999649 0.0000000 0.1667174 0.3333177
#
#       $exttime
#       [1] 1.0 2.0 2.5 3.0 4.0 5.0       (exttime include all the times,
#                                         censor or not, plus the added time)
#       $extstatus
```

```
#       [1]  1  0 -1  2  2  1
#
#       $extjump
#       [1] 0.4999649 0.0000000 0.1667174 0.0000000 0.0000000 0.3333177
#
#       $extsurv.Sx
#       [1] 0.5000351 0.5000351 0.3333177 0.3333177 0.3333177 0.0000000
#
#       $konstdist
#       [1] 0.4999365 0.4999365 0.6000000 0.6000000 0.6000000 1.0000000
#
#       $konstjump
#       [1] 0.4999365 0.0000000 0.1000635 0.0000000 0.0000000 0.4000000
#
#       $konsttime
#       [1] 3.5
#
#       $theta
#       [1] 0.6
#
#       $"-2loglikR"                  (the Wilks statistics to test Ho:
#       [1] 0.05679897                  F(K)=konst)
#
#       $maxiter
#       [1] 33
#
#  The Wilks statistic is only 0.05679897, there is no evidence agaist Ho
```

Index

∗Topic **nonparametric**
    `d011`, 1
    `d011ch`, 4
∗Topic **survival**
    `d011`, 1
    `d011ch`, 4

`d011`, 1
`d011ch`, 4