Package 'SCRSELECT'

August 23, 2017

Title Performs Bayesian Variable Selection on the Covariates in a Semi-Competing Risks Model

Version 1.3-3

Description Contains four functions used in the DIC-tau_g procedure. SCRSELECT() and SCRSELECTRUN() uses Stochastic Search Variable Selection to select important covariates in the three hazard functions of a semi-competing risks model. These functions perform the Gibbs sampler for variable selection and a Metropolis-Hastings-Green sampler for the number of split points and parameters for the three baseline hazard function. The function SCRSELECT() returns the posterior sample of all quantities sampled in the Gibbs sampler after a burn-in period to a desired file location, while the function SCRSELECTRUN() returns posterior values of important quantities to the DIC-Tau_g procedure in a list.
The function DICTAUG() returns a list containing the DIC values for the unique models visited by the DIC-Tau_g grid search.
The function ReturnModel() uses SCRSELECTRUN() and DICTAUG() to return a summary of the posterior coefficient vectors for the optimal model along with saving this posterior sample to a desired path location.

Depends R (>= 3.2.2), mvtnorm

License GPL-2

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Author Andrew Chapple [aut, cre]

Maintainer Andrew Chapple <Andrew.G.Chapple@rice.edu>

Repository CRAN

Date/Publication 2017-08-23 17:46:14 UTC

R topics documented:

DICTAUG *Performs a grid search over the marginal posterior probabilities of inclusion and returns a list of DIC values corresponding to each grid point. This is used in the ReturnModel function.*

Description

Performs a grid search over the marginal posterior probabilities of inclusion and returns a list of DIC values corresponding to each grid point. This is used in the ReturnModel function.

Usage

```
DICTAUG(PCT1, PCT2, PCT3, COV, Y1, Y2, I1, I2, s1, lam1, s2, lam2, s3, lam3,
  gam, c, B, inc)
```

Arguments

PCT1	Vector Containing posterior probabilities of inclusion for the hazard of a non-terminal event. This must be of length ncol(COV)-inc.
PCT2	Vector Containing posterior probabilities of inclusion for the hazard of death without a non-terminal event. This must be of length ncol(COV)-inc.
PCT3	Vector Containing posterior probabilities of inclusion for the hazard of death after a non-terminal event. This must be of length ncol(COV)-inc.
COV	Matrix of Patient Covariates. The last inc will be left out of variable selection.
Y1	Vector Containing non-terminal event times (or censoring time due to death/censoring).
Y2	Vector Containing Terminal Event times (or censoring).
I1	Vector Containing non-terminal event indicators (1 if non-terminal event for a patient, 0 otherwise).
I2	Vector Containing Terminal event indicators (1 if a patients experiences a non-ternminal event, 0 if censored).
s1	Vector containing the posterior locations of the split points in the hazard of a non-terminal event.
lam1	Vector containing the posterior log hazard heights on the split point intervals in the hazard of a non-terminal event.
s2	Vector containing the posterior locations of the split points in the hazard of death without a non-terminal event.
lam2	Vector containing the posterior log hazard heights on the split point intervals in the hazard of death without a non-terminal event.
s3	Vector containing the posterior locations of the split points in the hazard of death after a non-terminal event.
lam3	Vector containing the posterior log hazard heights on the split point intervals in the hazard of death after a non-terminal event.

gam	Vector of length n containing the posterior mean frailties of the patients.
c	Hyperparameter involved in the sampling of hazard coefficients. This should be the same value that controls the degree of sparsity achieved by the SVSS.
B	Number of iterations
inc	Number of variables left out of selection

Value

Returns a list of size 18 containing 18x18 matrices of DIC values and skipped entries.

@references [1] Lee, K. H., Haneuse, S., Schrag, D. and Dominici, F. (2015), Bayesian semi-parametric analysis of semi-competing risks data: investigating hospital readmission after a pancreatic cancer diagnosis. Journal of the Royal Statistical Society: Series C (Applied Statistics), 64: 253-273. doi: 10.1111/rssc.12078 [2] Chapple, A.C., Vannucci, M., Thall, P.F., Lin, S.(2017), Bayesian Variable selection for a semi-competing risks model with three hazard functions. Journal of Computational Statistics & Data Analysis, Volume 112, August 2017, Pages 170-185 [3] https://adventuresinstatistics.wordpress.com/2017/04/10/package-scrselect-using-returnmodel/

Examples

```
####Randomly Generate Semicompeting Risks Data
####Generates random patient time, indicator and covariates.
set.seed(1)
n=100
Y1=runif(n,0,100)
I1=rbinom(n,1,.5)
Y2=Y1
I2=I1
for(i in 1:n){if(I1[i]==0){Y2[i]=Y1[i]}else{Y2[i]=Y1[i]+runif(1,0,100)}}
I2=rbinom(n,1,.5)
library(mvtnorm)
X=rmvnorm(n,rep(0,7),diag(7))
###Read in Posterior mean quantities from SCRSELECTRETURN
PCT1=c(.2,.4,.7,.8,.5)
PCT2=c(.02,.06,.1,.5,.7)
PCT3=c(.85,.87,.3,.45,.51)
gam=rgamma(n,1,1)
s1=c(0,3,5,max(Y1[I1==1]))
lam1=c(-1,-3,0)
s2=c(0,1,max(Y2[I1==0]))
lam2=c(0,-2)
s3=c(0,max(Y2[I1==1]))
lam3=-2
####Read in Hyperparameters
c=5
###Number of iterations and output location
B=4
###Number of variables to exclude from selection and burnin percent
inc=2
DICTAUG(PCT1,PCT2,PCT3,X,Y1,Y2,I1,I2,s1,lam1,s2,lam2,s3,lam3,gam,c,B,inc)
```

`ReturnModel` *Performs the DIC-tau_g procedure and returns the posterior quantities of the optimal model.*

Description

Performs the DIC-tau_g procedure by first running the function SCRSELECTRUN with 60 percent burnin, which performs SVSS on two disperse starting values for beta1,beta2,beta3. Afterwards, the function DICTAUG is used to extract the DIC values for unique models visited by the grid search and the optimal model is determined as the one with the lowest DIC which is the most parsimonius. After the optimal model is determined, one final MCMC is performed to obtain posterior beta1,beta2 and beta3 quantities for this model, returning summary values for each hazard.

Usage

```
ReturnModel(Y1, I1, Y2, I2, X, hyperparameters, inc, c, BSVSS, BDIC, Path)
```

Arguments

- `Y1` Vector Containing non-terminal event times (or censoring time due to death/censoring)
- `I1` Vector Containing non-terminal event indicators (1 if non-terminal event for a patient, 0 otherwise)
- `Y2` Vector Containing Terminal Event times (or censoring)
- `I2` Vector Containing Terminal event indicators (1 if a patients experiences a nonternminal event, 0 if censored)
- `X` Matrix of Patient Covariates. The last inc will be left out of variable selection.
- `hyperparameters` List containing 29 hyperparameters and four starting values. In order they are: psi-the swap rate of the SVSS algorithm. c-parameter involved in Sigma matrix for selection. z1a, z1b, z2a, z2b, z3a, z3b - beta hyper parameters on probability of inclusion for each of the three hazard functions. a1,b1,a2,b2,a3,b3- hyperparameters on sigma_lambda_1, sigma_lambda_2, and sigma_lambda_3. clam1, clam2, clam3 - spatial dependency of baseline hazard (between 0 and 1) for the three hazard functions. Alpha1, Alpha2, Alpha3 - The parameter for the number of split points in hazards 1,2 and 3 (must be whole number). J1max, J2max, J3max - Maximum number of split points allowed (must be whole number). J1, J2, J3- Starting number of split points. w, psi1- hyperparameters on theta^-1. cep=Tuning Parameter for theta^-1 sampler. epstart-Starting value for theta^-1. cl1,cl2,cl3-Tuning parameters for log baseline hazard height sampler.
- `inc` Number of variables left out of selection.
- `c` sparsity parameter involved in Sigma matrix for selection. This should be the same c as that used in the hyperparameters vector.
- `BSVSS` Number of iterations to perform during the SVSS procedure. 100,000 is a reccomended value to achieve convergence.

BDIC	Number of iterations to perform during the DIC-tau_g grid search. 10,000 is a reccomended value to achieve convergence in a reasonable amount of time.
Path	Where to save posterior coefficient samples for the optimal model.

Value

Returns the optimal model determined by the DIC-Tau_g procedure and it's DIC along with summaries of these posterior quantities. Additionally, this function saves these posterior samples to a desired path.

@references [1] Lee, K. H., Haneuse, S., Schrag, D. and Dominici, F. (2015), Bayesian semiparametric analysis of semi-competing risks data: investigating hospital readmission after a pancreatic cancer diagnosis. Journal of the Royal Statistical Society: Series C (Applied Statistics), 64: 253-273. doi: 10.1111/rssc.12078 [2] Chapple, A.C., Vannucci, M., Thall, P.F., Lin, S.(2017), Bayesian Variable selection for a semi-competing risks model with three hazard functions. Journal of Computational Statistics & Data Analysis, Volume 112, August 2017, Pages 170-185 [3] https://adventuresinstatistics.wordpress.com/2017/04/10/package-scrselect-using-returnmodel/

Examples

```
####Randomly Generate Semicompeting Risks Data
set.seed(1)
####Generates random patient time, indicator and covariates.
n=100
Y1=runif(n,0,100)
I1=rbinom(n,1,.5)
Y2=Y1
I2=I1
for(i in 1:n){if(I1[i]==0){Y2[i]=Y1[i]}else{Y2[i]=Y1[i]+runif(1,0,100)}}
I2=rbinom(n,1,.5)
library(mvtnorm)
X=rmvnorm(n,rep(0,7),diag(7))
####Read in Hyperparameters
##Swap Rate
psi=.5
c=5
###Eta Beta function probabilities
z1a=.4
z1b=1.6
z2a=.4
z2b=1.6
z3a=.4
z3b=1.6
####Hierarchical lam params
###Sigma^2 lambda_g hyperparameters
a1=.7
b1=.7
a2=a1
b2=b1
a3=a1
b3=b1
##Spacing dependence c in [0,1]
```

```
clam1=1
clam2=1
clam3=1
#####NumSplit
alpha1=3
alpha2=3
alpha3=3
J1max=10
J2max=10
J3max=10
####Split Point Starting Value ###
J1=3
J2=3
J3=3
###epsilon starting values/hyperparameters###
w=.7
psi1=.7
cep=2.4
#############
epstart=1.5
cl1=.25
cl2=.25
cl3=.25
###Beta Starting Values
hyper1=c(psi,c,z1a,z1b,z2a,z2b,z3a,z3b,a1,b1,a2,b2,a3,b3,clam1,clam2,clam3)
hyper2=c(alpha1,alpha2,alpha3,J1max,J2max,J3max,J1,J2,J3,w,psi1,cep,epstart,cl1,cl2,cl3)
hyper=c(hyper1,hyper2)
###Number of iterations and output location
BSVSS=10
BDIC=4
Path=tempdir()
###Number of variables to exclude from selection and burnin percent
inc=2
ReturnModel(Y1,I1,Y2,I2,X,hyper,inc,c,BSVSS,BDIC,Path)
```

SCRSELECT *Performs Bayesian Variable Selection on the covariates in a semi-competing risks model*

Description

Performs Bayesian Variable Selection on the covariates in a semi-competing risks model

Usage

```
SCRSELECT(Y1, I1, Y2, I2, X, hyperparameters, beta1start, beta2start,
  beta3start, B, inc, Path, burn)
```

Arguments

Y1 Vector Containing non-terminal event times (or censoring time due to death/censoring)

I1 Vector Containing non-terminal event indicators (1 if non-terminal event for a patient, 0 otherwise)

Y2 Vector Containing Terminal Event times (or censoring)

I2 Vector Containing Terminal event indicators (1 if a patients experiences a non-ternminal event, 0 if censored)

X Matrix of Patient Covariates. The last inc will be left out of variable selection.

hyperparameters
List containing 29 hyperparameters and four starting values. In order they are: psi-the swap rate of the SVSS algorithm. c-parameter involved in Sigma matrix for selection. z1a, z1b, z2a, z2b, z3a, z3b - beta hyper parameters on probability of inclusion for each of the three hazard functions. a1,b1,a2,b2,a3,b3- hyperparameters on sigma_lambda_1, sigma_lambda_2, and sigma_lambda_3. clam1, clam2, clam3 - spatial dependency of baseline hazard (between 0 and 1) for the three hazard functions. Alpha1, Alpha2, Alpha3 - The parameter for the number of split points in hazards 1,2 and 3 (must be whole number). J1max, J2max, J3max - Maximum number of split points allowed (must be whole number). J1, J2, J3- Starting number of split points. w, psi1- hyperparameters on theta^-1. cep=Tuning Parameter for theta^-1 sampler. epstart-Starting value for theta^-1. cl1,cl2,cl3-Tuning parameters for log baseline hazard height sampler.

beta1start Starting Values for Beta1

beta2start Starting Values for Beta2

beta3start Starting Values for Beta3

B Number of iterations

inc Number of variables left out of selection

Path Where to save posterior samples

burn percent of posterior sample to burn in (burn*B must be a whole number)

Value

Returns marginal posterior probability of inclusion (post burn-in) for each hazard function along with acceptance rates for the various Metropolis-Hastings (and Metropolis-Hastings-Green) samplers.

References

[1] Lee, K. H., Haneuse, S., Schrag, D. and Dominici, F. (2015), Bayesian semi-parametric analysis of semi-competing risks data: investigating hospital readmission after a pancreatic cancer diagnosis. Journal of the Royal Statistical Society: Series C (Applied Statistics), 64: 253-273. doi: 10.1111/rssc.12078 [2] Chapple, A.C., Vannucci, M., Thall, P.F., Lin, S.(2017), Bayesian Variable selection for a semi-competing risks model with three hazard functions. Journal of Computational Statistics & Data Analysis, Volume 112, August 2017, Pages 170-185 [3] https://adventuresinstatistics.wordpress.com/2017/0
scrselect-using-returnmodel/

Examples

```
####Randomly Generate Semicompeting Risks Data
set.seed(1)
####Generates random patient time, indicator and covariates.
n=100
Y1=runif(n,0,100)
I1=rbinom(n,1,.5)
Y2=Y1
I2=I1
for(i in 1:n){if(I1[i]==0){Y2[i]=Y1[i]}else{Y2[i]=Y1[i]+runif(1,0,100)}}
I2=rbinom(n,1,.5)
library(mvtnorm)
X=rmvnorm(n,rep(0,7),diag(7))
####Read in Hyperparameters
##Swap Rate
psi=.5
c=5
###Eta Beta function probabilities
z1a=.4
z1b=1.6
z2a=.4
z2b=1.6
z3a=.4
z3b=1.6
####Hierarchical lam params
###Sigma^2 lambda_g hyperparameters
a1=.7
b1=.7
a2=a1
b2=b1
a3=a1
b3=b1
##Spacing dependence c in [0,1]
clam1=1
clam2=1
clam3=1
#####NumSplit
alpha1=3
alpha2=3
alpha3=3
J1max=10
J2max=10
J3max=10
####Split Point Starting Value ###
J1=3
J2=3
J3=3
###epsilon starting values/hyperparameters###
w=.7
psi1=.7
cep=2.4
#############
```

```
epstart=1.5
cl1=.25
cl2=.25
cl3=.25
###Beta Starting Values
beta1start=c(1,1,1,1,1,-1,-1)
beta2start=c(1,1,1,1,1,-1,-1)
beta3start=c(-1,1,1,1,1,-1,-1)
hyper1=c(psi,c,z1a,z1b,z2a,z2b,z3a,z3b,a1,b1,a2,b2,a3,b3,clam1,clam2,clam3)
hyper2=c(alpha1,alpha2,alpha3,J1max,J2max,J3max,J1,J2,J3,w,psi1,cep,epstart,cl1,cl2,cl3)
hyper=c(hyper1,hyper2)
###Number of iterations and output location
B=100
Path=tempdir()
###Number of variables to exclude from selection and burnin percent
inc=2
burn=.1
SCRSELECT(Y1,I1,Y2,I2,X,hyper,beta1start,beta2start,beta3start,B,inc,Path,burn)
```

`SCRSELECTRETURN` *Performs Bayesian Variable Selection on the covariates in a semi-competing risks model and returns burned in posterior means of parameters. This function is used in the ReturnModel function.*

Description

Performs Bayesian Variable Selection on the covariates in a semi-competing risks model and returns burned in posterior means of parameters. This function is used in the ReturnModel function.

Usage

```
SCRSELECTRETURN(Y1, I1, Y2, I2, X, hyperparameters, beta1start, beta2start,
  beta3start, B, inc, burn)
```

Arguments

`Y1` Vector Containing non-terminal event times (or censoring time due to death/censoring)

`I1` Vector Containing non-terminal event indicators (1 if non-terminal event for a patient, 0 otherwise)

`Y2` Vector Containing Terminal Event times (or censoring)

`I2` Vector Containing Terminal event indicators (1 if a patients experiences a non-ternminal event, 0 if censored)

`X` Matrix of Patient Covariates. The last inc will be left out of variable selection.

`hyperparameters`
List containing 29 hyperparameters and four starting values. In order they are: psi-the swap rate of the SVSS algorithm. c-parameter involved in Sigma matrix for selection. z1a, z1b, z2a, z2b, z3a, z3b - beta hyper parameters on probability

of inclusion for each of the three hazard functions. a1,b1,a2,b2,a3,b3- hyperparameters on sigma_lambda_1, sigma_lambda_2, and sigma_lambda_3. clam1, clam2, clam3 - spatial dependency of baseline hazard (between 0 and 1) for the three hazard functions. Alpha1, Alpha2, Alpha3 - The parameter for the number of split points in hazards 1,2 and 3 (must be whole number). J1max, J2max, J3max - Maximum number of split points allowed (must be whole number). J1, J2, J3- Starting number of split points. w, psi1- hyperparameters on theta^-1. cep=Tuning Parameter for theta^-1 sampler. epstart-Starting value for theta^-1. cl1,cl2,cl3-Tuning parameters for log baseline hazard height sampler.

`beta1start` Starting Values for Beta1

`beta2start` Starting Values for Beta2

`beta3start` Starting Values for Beta3

`B` Number of iterations

`inc` Number of variables left out of selection

`burn` percent of posterior sample to burn in (burn*B must be a whole number)

Value

Returns a list the following posterior quantities after burn in: Marginal probability of inclusion, mean frailty parameters, and the baseline hazard samples for each hazard.

References

[1] Lee, K. H., Haneuse, S., Schrag, D. and Dominici, F. (2015), Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. Journal of the Royal Statistical Society: Series C (Applied Statistics), 64: 253-273. doi: 10.1111/rssc.12078

Examples

```
####Randomly Generate Semicompeting Risks Data
set.seed(1)
####Generates random patient time, indicator and covariates.
n=100
Y1=runif(n,0,100)
I1=rbinom(n,1,.5)
Y2=Y1
I2=I1
for(i in 1:n){if(I1[i]==0){Y2[i]=Y1[i]}else{Y2[i]=Y1[i]+runif(1,0,100)}}
I2=rbinom(n,1,.5)
library(mvtnorm)
X=rmvnorm(n,rep(0,7),diag(7))
####Read in Hyperparameters
##Swap Rate
psi=.5
c=5
###Eta Beta function probabilities
z1a=.4
z1b=1.6
```

```
z2a=.4
z2b=1.6
z3a=.4
z3b=1.6
####Hierarchical lam params
###Sigma^2 lambda_g hyperparameters
a1=.7
b1=.7
a2=a1
b2=b1
a3=a1
b3=b1
##Spacing dependence c in [0,1]
clam1=1
clam2=1
clam3=1
#####NumSplit
alpha1=3
alpha2=3
alpha3=3
J1max=10
J2max=10
J3max=10
####Split Point Starting Value ###
J1=3
J2=3
J3=3
###epsilon starting values/hyperparameters###
w=.7
psi1=.7
cep=2.4
#############
epstart=1.5
cl1=.25
cl2=.25
cl3=.25
###Beta Starting Values
beta1start=c(1,1,1,1,1,-1,-1)
beta2start=c(1,1,1,1,1,-1,-1)
beta3start=c(-1,1,1,1,1,-1,-1)
hyper1=c(psi,c,z1a,z1b,z2a,z2b,z3a,z3b,a1,b1,a2,b2,a3,b3,clam1,clam2,clam3)
hyper2=c(alpha1,alpha2,alpha3,J1max,J2max,J3max,J1,J2,J3,w,psi1,cep,epstart,cl1,cl2,cl3)
hyper=c(hyper1,hyper2)
###Number of iterations and output location
B=100
Path=tempdir()
###Number of variables to exclude from selection and burnin percent
inc=2
burn=.1
SCRSELECTRETURN(Y1,I1,Y2,I2,X,hyper,beta1start,beta2start,beta3start,B,inc,burn)
```

Index