Package 'psbcGroup'

June 28, 2017

Type Package

Title Penalized Parametric and Semiparametric Bayesian Survival Models
with Shrinkage and Grouping Priors

Version 1.4

Date 2017-06-28

Author Kyu Ha Lee, Sounak Chakraborty, (Tony) Jianguo Sun

Maintainer Kyu Ha Lee <klee@hsph.harvard.edu>

Description Algorithms for fitting penalized parametric and semiparametric Bayesian survival models with shrinkage and grouping priors.

License GPL (>= 2)

Depends LearnBayes, SuppDists, mvtnorm, R (>= 3.2.3)

LazyLoad yes

NeedsCompilation yes

Repository CRAN

Date/Publication 2017-06-28 21:47:39 UTC

R topics documented:

1

| aftGL | *Function to Fit the Penalized Parametric Bayesian Accelerated Failure Time Model with Group Lasso Prior* |

Description

Penalized parametric Bayesian accelerated failure time model with group lasso prior is implemented to analyze survival data with high-dimensional covariates.

Usage

```
aftGL(survData, grpInx, hyperParams, startValues, numReps, thin, burninPerc = 0.5)
```

Arguments

survData
: a data.frame containing the time-to-event outcome, the censoring indicator, p covariate vectors from n subjects. It is of dimension $n \times (p + 2)$.

grpInx
: a vector of p group indicator for each variable

hyperParams
: a numeric vector containing hyperparameter values in hierarchical models: c(nu0, sigSq0, alpha0, h0, rLam, deltaLam). (nu0, sigSq0): hyperparameters for the prior of σ^2; (alpha0, h0): hyperparameters for the prior of α; (rLam, deltaLam): hyperparameters for the prior of λ^2.

startValues
: a numeric vector containing starting values for model parameters: c(alpha, beta, sigSq, tauSq, lambdaSq, w). See Examples below.

numReps
: total number of scans

thin
: extent of thinning

burninPerc
: the proportion of burn-in

Value

aftGL returns an object of class aftGL.

Author(s)

Kyu Ha Lee, Sounak Chakraborty, (Tony) Jianguo Sun

References

Lee, K. H., Chakraborty, S., and Sun, J. Variable Selection for High-Dimensional Genomic Data with Censored Outcomes Using Group Lasso Prior. *Computational Statistics and Data Analysis*, Volume 112, pages 1-13.

Examples

```
# generate some survival data
set.seed(204542)

p = 20
n = 100
logHR.true <- c(rep(4, 10), rep(0, (p-10)))

CovX<-matrix(0,p,p)

for(i in 1:10){
for(j in 1:10){
CovX[i,j] <- 0.5^abs(i-j)
}
}

diag(CovX) <- 1

x <- apply(rmvnorm(n, sigma=CovX, method="chol"), 2, scale)
pred <- as.vector(exp(rowSums(scale(x, center = FALSE, scale = 1/logHR.true))))

t  <- rexp(n, rate = pred)
cen <- runif(n, 0, 8)
tcen  <- pmin(t, cen)
di   <- as.numeric(t <= cen)

n <- dim(x)[1]
p <- dim(x)[2]

survData <- data.frame(cbind(tcen, di, x))
colnames(survData)[c(1:2)] <- c("time", "event")

grpInx <- 1:p
K <- length(unique(grpInx))

############################
nu0 <- 3
sigSq0 <- 1

alpha0 <- 0
h0 <- 10^6

rLam <- 0.5
deltaLam <- 2

hyperParams <- c(nu0, sigSq0, alpha0, h0, rLam, deltaLam)

############################
alpha <- 0.1
beta <- rep(0.1, p)
sigSq <- 1
```

```
tauSq <- rep(0.4, p)
lambdaSq <- 5
w <- log(tcen)

startValues <- as.vector(c(alpha, beta, sigSq, tauSq, lambdaSq, w))

############################
numReps <- 1000
thin <- 1
burninPerc = 0.5

fit <- aftGL(survData, grpInx, hyperParams, startValues, numReps, thin, burninPerc)
```

---

psbcEN   *Function to Fit the Penalized Semiparametric Bayesian Cox Model*
                                with Elastic Net Prior

Description

Penalized semiparametric Bayesian Cox (PSBC) model with elastic net prior is implemented to
analyze survival data with high-dimensional covariates.

Usage

```
psbcEN(survObj, priorPara, initial, rw=FALSE, mcmcPara, num.reps,
thin, chain = 1, save = 1000)
```

Arguments

survObj	The list containing observed data from n subjects; t, di, x
priorPara	The list containing prior parameter values; eta0, kappa0, c0, r1, r2, delta1, delta2, s
initial	The list containing the starting values of the parameters; beta.ini, lambda1Sq, lambda2, sigmaSq, tauSq, h
rw	When setting to "TRUE", the conventional random walk Metropolis Hastings algorithm is used. Otherwise, the mean and the variance of the proposal density is updated using the jumping rule described in Lee et al. (2011).
mcmcPara	The list containing the values of options for Metropolis-Hastings step for β; numBeta, beta.prop.var
num.reps	the number of iterations of the chain
thin	thinning
chain	the numeric name of chain in the case when running multiple chains.
save	frequency of storing the results in .Rdata file. For example, by setting "save = 1000", the algorithm saves the results every 1000 iterations.

Details

t	a vector of n times to the event
di	a vector of n censoring indicators for the event time (1=event occurred, 0=censored)
x	covariate matrix, n observations by p variables
eta0	scale parameter of gamma process prior for the cumulative baseline hazard, $eta0 > 0$
kappa0	shape parameter of gamma process prior for the cumulative baseline hazard, $kappa0 > 0$
c0	the confidence parameter of gamma process prior for the cumulative baseline hazard, $c0 > 0$
r1	the shape parameter of the gamma prior for λ_1^2
r2	the shape parameter of the gamma prior for λ_2
delta1	the rate parameter of the gamma prior for λ_1^2
delta2	the rate parameter of the gamma prior for λ_2
s	the set of time partitions for specification of the cumulative baseline hazard function
beta.ini	the starting values for β
lambda1Sq	the starting value for λ_1^2
lambda2	the starting value for λ_2
sigmaSq	the starting value for σ^2
tauSq	the starting values for τ^2
h	the starting values for h
numBeta	the number of components in β to be updated at one iteration
beta.prop.var	the variance of the proposal density for β when rw is set to "TRUE"

Value

psbcEN returns an object of class psbcEN

beta.p	posterior samples for β
h.p	posterior samples for h
tauSq.p	posterior samples for τ^2
mcmcOutcome	The list containing posterior samples for the remaining model parameters

Note

If the prespecified value of save is less than that of num.reps, the results are saved as .Rdata file under the directory working directory/mcmcOutcome.

Author(s)

Kyu Ha Lee, Sounak Chakraborty, (Tony) Jianguo Sun

References

Lee, K. H., Chakraborty, S., and Sun, J. (2011). Bayesian Variable Selection in Semiparametric Proportional Hazards Model for High Dimensional Survival Data. *The International Journal of Biostatistics*, Volume 7, Issue 1, Pages 1-32.

Lee, K. H., Chakraborty, S., and Sun, J. (2015). Survival Prediction and Variable Selection with Simultaneous Shrinkage and Grouping Priors. *Statistical Analysis and Data Mining*, Volume 8, Issue 2, pages 114-127.

Examples

```
## Not run:

# generate some survival data

set.seed(204542)

p = 20
n = 100
beta.true <- c(rep(4, 10), rep(0, (p-10)))

CovX<- diag(0.1, p)

survObj  <- list()
survObj$x <- apply(rmvnorm(n, sigma=CovX, method="chol"), 2, scale)

pred <- as.vector(exp(rowSums(scale(survObj$x, center = FALSE, scale = 1/beta.true))))

t  <- rexp(n, rate = pred)
cen <- runif(n, 0, 8)
survObj$t  <- pmin(t, cen)
survObj$di  <- as.numeric(t <= cen)

priorPara  <- list()
priorPara$eta0  <- 1
priorPara$kappa0  <- 1
priorPara$c0  <- 2
priorPara$r1 <- 0.1
priorPara$r2 <- 1
priorPara$delta1 <- 0.1
priorPara$delta2 <- 1
priorPara$s <- sort(survObj$t[survObj$di == 1])
priorPara$s <- c(priorPara$s, 2*max(survObj$t)
- max(survObj$t[-which(survObj$t==max(survObj$t))]))
priorPara$J <- length(priorPara$s)

mcmcPara <- list()
mcmcPara$numBeta <- p
mcmcPara$beta.prop.var <- 1

initial <- list()
initial$beta.ini <- rep(0.5, p)
initial$lambda1Sq <- 1
initial$lambda2 <- 1
initial$sigmaSq <- runif(1, 0.1, 10)
initial$tauSq <- rexp(p, rate = initial$lambda1Sq/2)
initial$h <- rgamma(priorPara$J, 1, 1)
```

```
rw = FALSE
num.reps = 20000
chain = 1
thin = 5
save = 5

fitEN <- psbcEN(survObj, priorPara, initial, rw=FALSE, mcmcPara,
num.reps, thin, chain, save)


## End(Not run)
```

---

psbcFL   *Function to Fit the Penalized Semiparametric Bayesian Cox Model*
                          with Fused Lasso Prior

Description

Penalized semiparametric Bayesian Cox (PSBC) model with fused lasso prior is implemented to analyze survival data with high-dimensional covariates.

Usage

```
psbcFL(survObj, priorPara, initial, rw=FALSE, mcmcPara, num.reps,
thin, chain = 1, save = 1000)
```

Arguments

survObj	The list containing observed data from n subjects; t, di, x
priorPara	The list containing prior parameter values; eta0, kappa0, c0, r1, r2, delta1, delta2, s
initial	The list containing the starting values of the parameters; beta.ini, lambda1Sq, lambda2Sq, sigmaSq, tauSq, h, wSq
rw	When setting to "TRUE", the conventional random walk Metropolis Hastings algorithm is used. Otherwise, the mean and the variance of the proposal density is updated using the jumping rule described in Lee et al. (2011).
mcmcPara	The list containing the values of options for Metropolis-Hastings step for β; numBeta, beta.prop.var
num.reps	the number of iterations of the chain
thin	thinning
chain	the numeric name of chain in the case when running multiple chains.
save	frequency of storing the results in .Rdata file. For example, by setting "save = 1000", the algorithm saves the results every 1000 iterations.

Details

t	a vector of n times to the event
di	a vector of n censoring indicators for the event time (1=event occurred, 0=censored)
x	covariate matrix, n observations by p variables
eta0	scale parameter of gamma process prior for the cumulative baseline hazard, $eta0 > 0$
kappa0	shape parameter of gamma process prior for the cumulative baseline hazard, $kappa0 > 0$
c0	the confidence parameter of gamma process prior for the cumulative baseline hazard, $c0 > 0$
r1	the shape parameter of the gamma prior for λ_1^2
r2	the shape parameter of the gamma prior for λ_2^2
delta1	the rate parameter of the gamma prior for λ_1^2
delta2	the rate parameter of the gamma prior for λ_2^2
s	the set of time partitions for specification of the cumulative baseline hazard function
beta.ini	the starting values for β
lambda1Sq	the starting value for λ_1^2
lambda2Sq	the starting value for λ_2^2
sigmaSq	the starting value for σ^2
tauSq	the starting values for τ^2
h	the starting values for h
wSq	the starting values for w^2
numBeta	the number of components in β to be updated at one iteration
beta.prop.var	the variance of the proposal density for β when rw is set to "TRUE"

Value

psbcFL returns an object of class psbcFL

beta.p	posterior samples for β
h.p	posterior samples for h
tauSq.p	posterior samples for τ^2
mcmcOutcome	The list containing posterior samples for the remaining model parameters

Note

If the prespecified value of save is less than that of num.reps, the results are saved as .Rdata file under the directory working directory/mcmcOutcome.

Author(s)

Kyu Ha Lee, Sounak Chakraborty, (Tony) Jianguo Sun

References

Lee, K. H., Chakraborty, S., and Sun, J. (2011). Bayesian Variable Selection in Semiparametric Proportional Hazards Model for High Dimensional Survival Data. *The International Journal of*

Biostatistics, Volume 7, Issue 1, Pages 1-32.

Lee, K. H., Chakraborty, S., and Sun, J. (2015). Survival Prediction and Variable Selection with Simultaneous Shrinkage and Grouping Priors. *Statistical Analysis and Data Mining*, Volume 8, Issue 2, pages 114-127.

Examples

```
## Not run:

# generate some survival data

set.seed(204542)

p = 20
n = 100
beta.true <- c(rep(4, 10), rep(0, (p-10)))

CovX<- diag(0.1, p)

survObj  <- list()
survObj$x <- apply(rmvnorm(n, sigma=CovX, method="chol"), 2, scale)

pred <- as.vector(exp(rowSums(scale(survObj$x, center = FALSE, scale = 1/beta.true))))

t   <- rexp(n, rate = pred)
cen <- runif(n, 0, 8)
survObj$t  <- pmin(t, cen)
survObj$di  <- as.numeric(t <= cen)

priorPara  <- list()
priorPara$eta0  <- 2
priorPara$kappa0  <- 2
priorPara$c0  <- 2
priorPara$r1 <- 0.5
priorPara$r2 <- 0.5
priorPara$delta1 <- 0.0001
priorPara$delta2 <- 0.0001
priorPara$s <- sort(survObj$t[survObj$di == 1])
priorPara$s <- c(priorPara$s, 2*max(survObj$t)
-max(survObj$t[-which(survObj$t==max(survObj$t))]))
priorPara$J <- length(priorPara$s)

mcmcPara <- list()
mcmcPara$numBeta <- p
mcmcPara$beta.prop.var <- 1

initial <- list()
initial$beta.ini <- rep(0.5, p)
initial$lambda1Sq <- 1
initial$lambda2Sq <- 1
```

```
initial$sigmaSq <- runif(1, 0.1, 10)
initial$tauSq <- rexp(p, rate = initial$lambda1Sq/2)
initial$h <- rgamma(priorPara$J, 1, 1)
initial$wSq   <- rexp((p-1), rate = initial$lambda2Sq/2)

rw = FALSE
num.reps = 20000
chain = 1
thin = 5
save = 5

fitFL <- psbcFL(survObj, priorPara, initial, rw=FALSE, mcmcPara,
num.reps, thin, chain, save)

## End(Not run)
```

---

psbcGL   *Function to Fit the Penalized Semiparametric Bayesian Cox Model*
                            with Group Lasso Prior

Description

Penalized semiparametric Bayesian Cox (PSBC) model with group lasso prior is implemented to
analyze survival data with high-dimensional covariates.

Usage

```
psbcGL(survObj, priorPara, initial, rw=FALSE, mcmcPara, num.reps,
thin, chain = 1, save = 1000)
```

Arguments

survObj	The list containing observed data from n subjects; t, di, x
priorPara	The list containing prior parameter values; eta0, kappa0, c0, r, delta, s, groupInd
initial	The list containing the starting values of the parameters; beta.ini, lambdaSq, sigmaSq, tauSq, h
rw	When setting to "TRUE", the conventional random walk Metropolis Hastings algorithm is used. Otherwise, the mean and the variance of the proposal density is updated using the jumping rule described in Lee et al. (2011).
mcmcPara	The list containing the values of options for Metropolis-Hastings step for β; numBeta, beta.prop.var
num.reps	the number of iterations of the chain
thin	thinning
chain	the numeric name of chain in the case when running multiple chains.
save	frequency of storing the results in .Rdata file. For example, by setting "save = 1000", the algorithm saves the results every 1000 iterations.

Details

t	a vector of n times to the event
di	a vector of n censoring indicators for the event time (1=event occurred, 0=censored)
x	covariate matrix, n observations by p variables
eta0	scale parameter of gamma process prior for the cumulative baseline hazard, $eta0 > 0$
kappa0	shape parameter of gamma process prior for the cumulative baseline hazard, $kappa0 > 0$
c0	the confidence parameter of gamma process prior for the cumulative baseline hazard, $c0 > 0$
r	the shape parameter of the gamma prior for λ^2
delta	the rate parameter of the gamma prior for λ^2
s	the set of time partitions for specification of the cumulative baseline hazard function
groupInd	a vector of p group indicator for each variable
beta.ini	the starting values for β
lambdaSq	the starting value for λ^2
sigmaSq	the starting value for σ^2
tauSq	the starting values for τ^2
h	the starting values for h
numBeta	the number of components in β to be updated at one iteration
beta.prop.var	the variance of the proposal density for β when rw is set to "TRUE"

Value

psbcGL returns an object of class psbcGL

beta.p	posterior samples for β
h.p	posterior samples for h
tauSq.p	posterior samples for τ^2
mcmcOutcome	The list containing posterior samples for the remaining model parameters

Note

To fit the PSBC model with the ordinary Bayesian lasso prior (Lee et al., 2011), groupInd needs to be set to 1:p. If the prespecified value of save is less than that of num.reps, the results are saved as .Rdata file under the directory working directory/mcmcOutcome.

Author(s)

Kyu Ha Lee, Sounak Chakraborty, (Tony) Jianguo Sun

References

Lee, K. H., Chakraborty, S., and Sun, J. (2011). Bayesian Variable Selection in Semiparametric Proportional Hazards Model for High Dimensional Survival Data. *The International Journal of Biostatistics*, Volume 7, Issue 1, Pages 1-32.

Lee, K. H., Chakraborty, S., and Sun, J. (2015). Survival Prediction and Variable Selection with Simultaneous Shrinkage and Grouping Priors. *Statistical Analysis and Data Mining*, Volume 8, Issue 2, pages 114-127.

Examples

```
## Not run:

# generate some survival data

set.seed(204542)

p = 20
n = 100
beta.true <- c(rep(4, 10), rep(0, (p-10)))

CovX<-matrix(0,p,p)

for(i in 1:10){
for(j in 1:10){
CovX[i,j] <- 0.5^abs(i-j)
}
}

diag(CovX) <- 1

survObj   <- list()
survObj$x <- apply(rmvnorm(n, sigma=CovX, method="chol"), 2, scale)

pred <- as.vector(exp(rowSums(scale(survObj$x, center = FALSE, scale = 1/beta.true))))

t   <- rexp(n, rate = pred)
cen <- runif(n, 0, 8)
survObj$t   <- pmin(t, cen)
survObj$di  <- as.numeric(t <= cen)

priorPara   <- list()
priorPara$eta0   <- 1
priorPara$kappa0   <- 1
priorPara$c0   <- 2
priorPara$r <- 0.5
priorPara$delta <- 0.0001
priorPara$s <- sort(survObj$t[survObj$di == 1])
priorPara$s <- c(priorPara$s, 2*max(survObj$t)
-max(survObj$t[-which(survObj$t==max(survObj$t))]))
priorPara$J <- length(priorPara$s)
priorPara$groupInd <- c(rep(1,10),2:11)

mcmcPara <- list()
mcmcPara$numBeta <- p
mcmcPara$beta.prop.var <- 1
```

```
initial <- list()
initial$beta.ini <- rep(0.5, p)
initial$lambdaSq <- 1
initial$sigmaSq <- runif(1, 0.1, 10)
initial$tauSq <- rexp(length(unique(priorPara$groupInd)),
rate = initial$lambdaSq/2)
initial$h <- rgamma(priorPara$J, 1, 1)

rw = FALSE
num.reps = 20000
chain = 1
thin = 5
save = 5

fitGL <- psbcGL(survObj, priorPara, initial, rw=FALSE, mcmcPara,
num.reps, thin, chain, save)


## End(Not run)
```

---

psbcGroup   *Penalized Parametric and Semiparametric Bayesian Survival Models*
                   with Shrinkage and Grouping Priors

Description

The package provides algorithms for fitting penalized parametric and semiparametric Bayesian survival models with elastic net, fused lasso, and group lasso priors.

Details

The package includes following functions:

psbcEN	The function to fit the PSBC model with elastic net prior
psbcFL	The function to fit the PSBC model with fused lasso prior
psbcGL	The function to fit the PSBC model with group lasso or Bayesian lasso prior
aftGL	The function to fit the parametric accelerated failure time model with group lasso

Package:	psbcGroup
Type:	Package
Version:	1.4
Date:	2017-06-28
License:	GPL (>= 2)
LazyLoad:	yes

Author(s)

Kyu Ha Lee, Sounak Chakraborty, (Tony) Jianguo Sun
Maintainer: Kyu Ha Lee <klee@hsph.harvard.edu>

References

Lee, K. H., Chakraborty, S., and Sun, J. (2011). Bayesian Variable Selection in Semiparametric Proportional Hazards Model for High Dimensional Survival Data. *The International Journal of Biostatistics*, Volume 7, Issue 1, Pages 1-32.

Lee, K. H., Chakraborty, S., and Sun, J. (2015). Survival Prediction and Variable Selection with Simultaneous Shrinkage and Grouping Priors. *Statistical Analysis and Data Mining*, Volume 8, Issue 2, pages 114-127.

Lee, K. H., Chakraborty, S., and Sun, J. Variable Selection for High-Dimensional Genomic Data with Censored Outcomes Using Group Lasso Prior. *Computational Statistics and Data Analysis*, Volume 112, pages 1-13.

Index