

Package ‘horseshoenlm’

December 18, 2020

Title Nonlinear Regression using Horseshoe Prior

Version 0.0.6

Description

Provides the posterior estimates of the regression coefficients when horseshoe prior is specified. The regression models considered here are logistic model for binary response and log normal accelerated failure time model for right censored survival response. The linear model analysis is also available for completeness. All models provide deviance information criterion and widely applicable information criterion. See <doi:10.1111/rssc.12377> Maity et. al. (2019) <doi:10.1111/biom.13132> Maity et. al. (2020).

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

Imports survival, msm

RoxygenNote 7.1.1

Suggests boot, pgdraw, mvtnorm

NeedsCompilation no

Author Arnab Kumar Maity [aut, cre]

Maintainer Arnab Kumar Maity <Arnab.Maity@pfizer.com>

Repository CRAN

Date/Publication 2020-12-18 10:10:02 UTC

R topics documented:

afths	2
lmhs	5
logiths	7
probiths	10
Index	13

Description

This function employs the algorithm provided by van der Pas et. al. (2016) for log normal Accelerated Failure Rate (AFT) model to fit survival regression. The censored observations are updated according to the data augmentation approach described in Maity et. al. (2019) and Maity et. al. (2020).

Usage

```
afths(
  ct,
  X,
  method.tau = c("fixed", "truncatedCauchy", "halfCauchy"),
  tau = 1,
  method.sigma = c("fixed", "Jeffreys"),
  Sigma2 = 1,
  burn = 1000,
  nmc = 5000,
  thin = 1,
  alpha = 0.05,
  Xtest = NULL
)
```

Arguments

ct	survival response, a $n \times 2$ matrix with first column as response and second column as right censored indicator, 1 is event time and 0 is right censored.
X	Matrix of covariates, dimension $n \times p$.
method.tau	Method for handling τ . Select "truncatedCauchy" for full Bayes with the Cauchy prior truncated to $[1/p, 1]$, "halfCauchy" for full Bayes with the half-Cauchy prior, or "fixed" to use a fixed value (an empirical Bayes estimate, for example).
tau	Use this argument to pass the (estimated) value of τ in case "fixed" is selected for method.tau. Not necessary when method.tau is equal to "halfCauchy" or "truncatedCauchy". The default (tau = 1) is not suitable for most purposes and should be replaced.
method.sigma	Select "Jeffreys" for full Bayes with Jeffrey's prior on the error variance σ^2 , or "fixed" to use a fixed value (an empirical Bayes estimate, for example).
Sigma2	A fixed value for the error variance σ^2 . Not necessary when method.sigma is equal to "Jeffreys". Use this argument to pass the (estimated) value of Sigma2 in case "fixed" is selected for method.sigma. The default (Sigma2 = 1) is not suitable for most purposes and should be replaced.
burn	Number of burn-in MCMC samples. Default is 1000.

nmc	Number of posterior draws to be saved. Default is 5000.
thin	Thinning parameter of the chain. Default is 1 (no thinning).
alpha	Level for the credible intervals. For example, alpha = 0.05 results in 95% credible intervals.
Xtest	test design matrix.

Details

The model is: t_i is response, c_i is censored time, $t_i^* = \min(t_i, c_i)$ is observed time, w_i is censored data, so $w_i = \log t_i^*$ if t_i is event time and $w_i = \log t_i^*$ if t_i is right censored $\log t_i = X\beta + \epsilon$, $\epsilon \sim N(0, \sigma^2)$.

Value

SurvivalHat	Predictive survival probability
LogTimeHat	Predictive log time
BetaHat	Posterior mean of Beta, a p by 1 vector
LeftCI	The left bounds of the credible intervals
RightCI	The right bounds of the credible intervals
BetaMedian	Posterior median of Beta, a p by 1 vector
LambdaHat	Posterior samples of λ , a $p * 1$ vector
Sigma2Hat	Posterior mean of error variance σ^2 . If method.sigma = "fixed" is used, this value will be equal to the user-selected value of Sigma2 passed to the function
TauHat	Posterior mean of global scale parameter tau, a positive scalar
BetaSamples	Posterior samples of β
TauSamples	Posterior samples of τ
Sigma2Samples	Posterior samples of Sigma2
LikelihoodSamples	Posterior samples of likelihood
DIC	Devainace Information Criterion of the fitted model
WAIC	Widely Applicable Information Criterion

References

- Maity, A. K., Carroll, R. J., and Mallick, B. K. (2019) "Integration of Survival and Binary Data for Variable Selection and Prediction: A Bayesian Approach", *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.
- Maity, A. K., Bhattacharya, A., Mallick, B. K., & Baladandayuthapani, V. (2020). Bayesian data integration and variable selection for pan cancer survival prediction using protein expression data. *Biometrics*, 76(1), 316-325.
- Stephanie van der Pas, James Scott, Antik Chakraborty and Anirban Bhattacharya (2016). horse-shoe: Implementation of the Horseshoe Prior. R package version 0.1.0. <https://CRAN.R-project.org/package=horseshoe>
- Enes Makalic and Daniel Schmidt (2016). High-Dimensional Bayesian Regularised Regression with the BayesReg Package arXiv:1611.06649

Examples

```

burnin <- 500
nmc <- 1000
thin <- 1
y.sd <- 1 # standard deviation of the response

p <- 100 # number of predictors
ntrain <- 100 # training size
ntest <- 50 # test size
n <- ntest + ntrain # sample size
q <- 10 # number of true predictors

beta.t <- c(sample(x = c(1, -1), size = q, replace = TRUE), rep(0, p - q))
x <- mvtnorm::rmvnorm(n, mean = rep(0, p), sigma = diag(p))

tmean <- x %*% beta.t
y <- rnorm(n, mean = tmean, sd = y.sd)
X <- scale(as.matrix(x)) # standarization

T <- exp(y) # AFT model
C <- rgamma(n, shape = 1.75, scale = 3) # 42% censoring time
time <- pmin(T, C) # observed time is min of censored and true
status = time == T # set to 1 if event is observed
ct <- as.matrix(cbind(time = time, status = status)) # censored time

# Training set
cttrain <- ct[1:ntrain, ]
Xtrain <- X[1:ntrain, ]

# Test set
cttest <- ct[(ntrain + 1):n, ]
Xtest <- X[(ntrain + 1):n, ]

posterior.fit <- afths(ct = cttrain, X = Xtrain, method.tau = "halfCauchy",
                      method.sigma = "Jeffreys", burn = burnin, nmc = nmc, thin = 1,
                      Xtest = Xtest)

posterior.fit$BetaHat

# Posterior processing to recover the true predictors
cluster <- kmeans(abs(posterior.fit$BetaHat), centers = 2)$cluster
cluster1 <- which(cluster == 1)
cluster2 <- which(cluster == 2)
min.cluster <- ifelse(length(cluster1) < length(cluster2), 1, 2)
which(cluster == min.cluster) # this matches with the true variables

```

lmhs

*Horseshoe shrinkage prior in Bayesian linear regression***Description**

This function employs the algorithm provided by van der Pas et. al. (2016) for linear model to fit Bayesian regression.

Usage

```
lmhs(
  y,
  X,
  method.tau = c("fixed", "truncatedCauchy", "halfCauchy"),
  tau = 1,
  method.sigma = c("fixed", "Jeffreys"),
  Sigma2 = 1,
  burn = 1000,
  nmc = 5000,
  thin = 1,
  alpha = 0.05,
  Xtest = NULL
)
```

Arguments

y	Response vector.
X	Matrix of covariates, dimension $n * p$.
method.tau	Method for handling τ . Select "truncatedCauchy" for full Bayes with the Cauchy prior truncated to $[1/p, 1]$, "halfCauchy" for full Bayes with the half-Cauchy prior, or "fixed" to use a fixed value (an empirical Bayes estimate, for example).
tau	Use this argument to pass the (estimated) value of τ in case "fixed" is selected for method.tau. Not necessary when method.tau is equal to "halfCauchy" or "truncatedCauchy". The default (tau = 1) is not suitable for most purposes and should be replaced.
method.sigma	Select "Jeffreys" for full Bayes with Jeffrey's prior on the error variance σ^2 , or "fixed" to use a fixed value (an empirical Bayes estimate, for example).
Sigma2	A fixed value for the error variance σ^2 . Not necessary when method.sigma is equal to "Jeffreys". Use this argument to pass the (estimated) value of Sigma2 in case "fixed" is selected for method.sigma. The default (Sigma2 = 1) is not suitable for most purposes and should be replaced.
burn	Number of burn-in MCMC samples. Default is 1000.
nmc	Number of posterior draws to be saved. Default is 5000.
thin	Thinning parameter of the chain. Default is 1 (no thinning).

alpha	Level for the credible intervals. For example, alpha = 0.05 results in 95% credible intervals.
Xtest	test design matrix.

Details

The model is: y_i is response, $y_i = X\beta + \epsilon$, $\epsilon \sim N(0, \sigma^2)$.

Value

yHat	Predictive response
BetaHat	Posterior mean of Beta, a p by 1 vector
LeftCI	The left bounds of the credible intervals
RightCI	The right bounds of the credible intervals
BetaMedian	Posterior median of Beta, a p by 1 vector
LambdaHat	Posterior samples of λ , a $p * 1$ vector
Sigma2Hat	Posterior mean of error variance σ^2 . If method.sigma = "fixed" is used, this value will be equal to the user-selected value of Sigma2 passed to the function
TauHat	Posterior mean of global scale parameter tau, a positive scalar
BetaSamples	Posterior samples of β
TauSamples	Posterior samples of τ
Sigma2Samples	Posterior samples of Sigma2
LikelihoodSamples	Posterior samples of likelihood
DIC	Devainace Information Criterion of the fitted model
WAIC	Widely Applicable Information Criterion

References

Maity, A. K., Carroll, R. J., and Mallick, B. K. (2019) "Integration of Survival and Binary Data for Variable Selection and Prediction: A Bayesian Approach", Journal of the Royal Statistical Society: Series C (Applied Statistics).

Maity, A. K., Bhattacharya, A., Mallick, B. K., & Baladandayuthapani, V. (2020). Bayesian data integration and variable selection for pan cancer survival prediction using protein expression data. *Biometrics*, 76(1), 316-325.

Stephanie van der Pas, James Scott, Antik Chakraborty and Anirban Bhattacharya (2016). horseshoe: Implementation of the Horseshoe Prior. R package version 0.1.0. <https://CRAN.R-project.org/package=horseshoe>

Enes Makalic and Daniel Schmidt (2016). High-Dimensional Bayesian Regularised Regression with the BayesReg Package arXiv:1611.06649

Examples

```

burnin <- 500
nmc <- 1000
thin <- 1
y.sd <- 1 # standard deviation of the response

p <- 100 # number of predictors
ntrain <- 100 # training size
ntest <- 50 # test size
n <- ntest + ntrain # sample size
q <- 10 # number of true predictors

beta.t <- c(sample(x = c(1, -1), size = q, replace = TRUE), rep(0, p - q))
x <- mvtnorm::rmvnorm(n, mean = rep(0, p), sigma = diag(p))

tmean <- x %*% beta.t
y <- rnorm(n, mean = tmean, sd = y.sd)
X <- scale(as.matrix(x)) # standarization

# Training set
ytrain <- y[1:ntrain]
Xtrain <- X[1:ntrain, ]

# Test set
ytest <- y[(ntrain + 1):n]
Xtest <- X[(ntrain + 1):n, ]

posterior.fit <- lmhs(y = ytrain, X = Xtrain, method.tau = "halfCauchy",
                    method.sigma = "Jeffreys", burn = burnin, nmc = nmc, thin = 1,
                    Xtest = Xtest)

posterior.fit$BetaHat

# Posterior processing to recover the true predictors
cluster <- kmeans(abs(posterior.fit$BetaHat), centers = 2)$cluster
cluster1 <- which(cluster == 1)
cluster2 <- which(cluster == 2)
min.cluster <- ifelse(length(cluster1) < length(cluster2), 1, 2)
which(cluster == min.cluster) # this matches with the true variables

```

Description

This function employs the algorithm provided by Makalic and Schmidt (2016) for binary logistic model to fit Bayesian logistic regression. The observations are updated according to the Polya-Gamma data augmentation of approach of Polson, Scott, and Windle (2014).

Usage

```
logiths(
  z,
  X,
  method.tau = c("fixed", "truncatedCauchy", "halfCauchy"),
  tau = 1,
  burn = 1000,
  nmc = 5000,
  thin = 1,
  alpha = 0.05,
  Xtest = NULL
)
```

Arguments

<code>z</code>	Response, a $n * 1$ vector of 1 or 0.
<code>X</code>	Matrix of covariates, dimension $n * p$.
<code>method.tau</code>	Method for handling τ . Select "truncatedCauchy" for full Bayes with the Cauchy prior truncated to $[1/p, 1]$, "halfCauchy" for full Bayes with the half-Cauchy prior, or "fixed" to use a fixed value (an empirical Bayes estimate, for example).
<code>tau</code>	Use this argument to pass the (estimated) value of τ in case "fixed" is selected for <code>method.tau</code> . Not necessary when <code>method.tau</code> is equal to "halfCauchy" or "truncatedCauchy". The default (<code>tau = 1</code>) is not suitable for most purposes and should be replaced.
<code>burn</code>	Number of burn-in MCMC samples. Default is 1000.
<code>nmc</code>	Number of posterior draws to be saved. Default is 5000.
<code>thin</code>	Thinning parameter of the chain. Default is 1 (no thinning).
<code>alpha</code>	Level for the credible intervals. For example, <code>alpha = 0.05</code> results in 95% credible intervals.
<code>Xtest</code>	test design matrix.

Details

The model is: z_i is response either 1 or 0, $\log \Pr(z_i = 1) = \text{logit}^{-1}(X\beta)$.

Value

<code>ProbHat</code>	Predictive probability
<code>BetaHat</code>	Posterior mean of Beta, a p by 1 vector
<code>LeftCI</code>	The left bounds of the credible intervals
<code>RightCI</code>	The right bounds of the credible intervals
<code>BetaMedian</code>	Posterior median of Beta, a p by 1 vector
<code>LambdaHat</code>	Posterior samples of λ , a $p * 1$ vector
<code>TauHat</code>	Posterior mean of global scale parameter tau, a positive scalar

BetaSamples	Posterior samples of β
TauSamples	Posterior samples of τ
LikelihoodSamples	Posterior samples of likelihood
DIC	Devainace Information Criterion of the fitted model
WAIC	Widely Applicable Information Criterion

References

- Stephanie van der Pas, James Scott, Antik Chakraborty and Anirban Bhattacharya (2016). horse-shoe: Implementation of the Horseshoe Prior. R package version 0.1.0. <https://CRAN.R-project.org/package=horseshoe>
- Enes Makalic and Daniel Schmidt (2016). High-Dimensional Bayesian Regularised Regression with the BayesReg Package arXiv:1611.06649
- Polson, N.G., Scott, J.G. and Windle, J. (2014) The Bayesian Bridge. Journal of Royal Statistical Society, B, 76(4), 713-733.

Examples

```

burnin <- 100
nmc <- 500
thin <- 1

p <- 100 # number of predictors
ntrain <- 250 # training size
ntest <- 100 # test size
n <- ntest + ntrain # sample size
q <- 10 # number of true predictos

beta.t <- c(sample(x = c(1, -1), size = q, replace = TRUE), rep(0, p - q))
x <- mvtnorm::rmvnorm(n, mean = rep(0, p), sigma = diag(p))

zmean <- x %*% beta.t
z <- rbinom(n, size = 1, prob = boot::inv.logit(zmean))
X <- scale(as.matrix(x)) # standarization

# Training set
ztrain <- z[1:ntrain]
Xtrain <- X[1:ntrain, ]

# Test set
ztest <- z[(ntrain + 1):n]
Xtest <- X[(ntrain + 1):n, ]

posterior.fit <- logiths(z = ztrain, X = Xtrain, method.tau = "halfCauchy",
                        burn = burnin, nmc = nmc, thin = 1,
                        Xtest = Xtest)

```

```

posterior.fit$BetaHat

# Posterior processing to recover the true predictors
cluster <- kmeans(abs(posterior.fit$BetaHat), centers = 2)$cluster
cluster1 <- which(cluster == 1)
cluster2 <- which(cluster == 2)
min.cluster <- ifelse(length(cluster1) < length(cluster2), 1, 2)
which(cluster == min.cluster) # this matches with the true variables

```

probiths

Horseshoe shrinkage prior in Bayesian Probit regression

Description

This function employs the algorithm provided by Makalic and Schmidt (2016) for binary probit model to fit Bayesian probit regression. The observations are updated according to the data augmentation of approach of Albert and Chib (1993).

The model is: z_i is response either 1 or 0, $\log \Pr(z_i = 1) = \Phi(X\beta)$, $\Phi \sim N(0, \sigma^2)$.

Usage

```

probiths(
  z,
  X,
  method.tau = c("fixed", "truncatedCauchy", "halfCauchy"),
  tau = 1,
  burn = 1000,
  nmc = 5000,
  thin = 1,
  alpha = 0.05,
  Xtest = NULL
)

```

Arguments

<code>z</code>	Response, a $n * 1$ vector of 1 or 0.
<code>X</code>	Matrix of covariates, dimension $n * p$.
<code>method.tau</code>	Method for handling τ . Select "truncatedCauchy" for full Bayes with the Cauchy prior truncated to $[1/p, 1]$, "halfCauchy" for full Bayes with the half-Cauchy prior, or "fixed" to use a fixed value (an empirical Bayes estimate, for example).
<code>tau</code>	Use this argument to pass the (estimated) value of τ in case "fixed" is selected for method.tau. Not necessary when method.tau is equal to "halfCauchy" or "truncatedCauchy". The default (tau = 1) is not suitable for most purposes and should be replaced.

burn	Number of burn-in MCMC samples. Default is 1000.
nmc	Number of posterior draws to be saved. Default is 5000.
thin	Thinning parameter of the chain. Default is 1 (no thinning).
alpha	Level for the credible intervals. For example, alpha = 0.05 results in 95% credible intervals.
Xtest	test design matrix.

Value

ProbHat	Predictive probability
BetaHat	Posterior mean of Beta, a p by 1 vector
LeftCI	The left bounds of the credible intervals
RightCI	The right bounds of the credible intervals
BetaMedian	Posterior median of Beta, a p by 1 vector
LambdaHat	Posterior samples of λ , a $p * 1$ vector
TauHat	Posterior mean of global scale parameter tau, a positive scalar
BetaSamples	Posterior samples of β
TauSamples	Posterior samples of τ
LikelihoodSamples	Posterior samples of likelihood
DIC	Devainace Information Criterion of the fitted model
WAIC	Widely Applicable Information Criterion

References

- Stephanie van der Pas, James Scott, Antik Chakraborty and Anirban Bhattacharya (2016). horse-shoe: Implementation of the Horseshoe Prior. R package version 0.1.0. <https://CRAN.R-project.org/package=horseshoe>
- Enes Makalic and Daniel Schmidt (2016). High-Dimensional Bayesian Regularised Regression with the BayesReg Package arXiv:1611.06649
- Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422), 669-679.

Examples

```
burnin <- 100
nmc <- 200
thin <- 1
y.sd <- 1 # statndard deviation of the response

p <- 200 # number of predictors
ntrain <- 250 # training size
ntest <- 100 # test size
n <- ntest + ntrain # sample size
q <- 10 # number of true predictos
```

```
beta.t <- c(sample(x = c(1, -1), size = q, replace = TRUE), rep(0, p - q))
x <- mvtnorm::rmvnorm(n, mean = rep(0, p))
zmean <- x %*% beta.t

y <- rnorm(n, mean = zmean, sd = y.sd)
z <- ifelse(y > 0, 1, 0)
X <- scale(as.matrix(x)) # standarization
z <- as.numeric(as.matrix(c(z)))

# Training set
ztrain <- z[1:ntrain]
Xtrain <- X[1:ntrain, ]

# Test set
ztest <- z[(ntrain + 1):n]
Xtest <- X[(ntrain + 1):n, ]

posterior.fit <- probiths(z = ztrain, X = Xtrain, method.tau = "halfCauchy",
                        burn = burnin, nmc = nmc, thin = 1,
                        Xtest = Xtest)

posterior.fit$BetaHat

# Posterior processing to recover the significant predictors
cluster <- kmeans(abs(posterior.fit$BetaHat), centers = 2)$cluster # return cluster indices
cluster1 <- which(cluster == 1)
cluster2 <- which(cluster == 2)
min.cluster <- ifelse(length(cluster1) < length(cluster2), 1, 2)
which(cluster == min.cluster) # this matches with the true variables
```

Index

afths, [2](#)

lmhs, [5](#)

logiths, [7](#)

probiths, [10](#)