Package 'BayesPen'

July 2, 2014

Type Package

Title Bayesian Penalized Credible Regions

Version 1.0

Date 2014-02-02

Author Ander Wilson, Howard D. Bondell, and Brian J. Reich

Maintainer Ander Wilson <ander_wilson@ncsu.edu>

Description Fits the Bayesian penalized credible region variable and confounder selection methods.

License GPL (>= 2)

Depends lars, MCMCpack, SuppDists

NeedsCompilation yes

Repository CRAN

Date/Publication 2014-04-29 22:05:44

R topics documented:

BayesPen-package *Bayesian Penalized Credible Regions*

Description

Performs the Bayesian penalized credible regions methods for variable selection (Bondell and Reich 2012) and confounder selection (Wilson and Reich 2014+). The solution path can be computed from the posterior mean and covariance matrix for a variety of models. The function BayesPen performs confounder or variable selection from the posterior means and covariances of a Bayesian regression model. The wrapper functions BayesPen.lm and BayesPen.lm.confounders fit linear models with code based on the BLM package (Campos and Rodriguez 2012) and run BayesPen together in one step.

Details

Package:	BayesPen
Type:	Package
Version:	1.0
Date:	2014-02-16
License:	GPL (>= 2)

Author(s)

Ander Wilson, Howard D. Bondell, and Brian J. Reich

Maintainer: Ander Wilson <ander_wilson@ncsu.edu>

References

Bondell, H. D. and Reich, B. J. (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. *J. Am. Statist. Assoc*. 107, 1610-1624.

Gustavo de los Campos and Paulino Perez Rodriguez, (2012). BLR: Bayesian Linear Regression. R package version 1.3. http://CRAN.R-project.org/package=BLR

Wilson, A. and Reich, B.J. (2014+). Confounder selection via penalized credible regions.

Examples

```
######################
#Variable Selection
set.seed(1234)
dat <- SimExample(500,model="BR1")
X <- dat$X
y <- dat$y
```

```
#fit the full model assuming flat priors on beta
fit1 <- lm(y~X-1)
betahat <- coef(fit1)
cov <- vcov(fit1)

#find solution path
fit.BayesPen <- BayesPen(beta=betahat, beta_cov=cov)

#refit the model
refit <- BayesPen.refit(y,X,fit.BayesPen)

#plot it
BayesPen.plot(refit)


######################
#Confounder Selection
set.seed(1234)
dat <- SimExample(500,model="WPD2")
X <- dat$X
U <- dat$U
W <- cbind(X,U)
y <- dat$y

#fit the full outcome model assuming flat priors on beta
fit1 <- lm(y~W-1)
betahat <- coef(fit1)
cov <- vcov(fit1)

#fit the full exposure model assuming flat priors on beta
fit2 <- lm(X~U-1)
gammahat <- coef(fit2)

#find solution path
fit.BayesPen <- BayesPen(beta=betahat, beta_cov=cov, confounder.weights=c(0,gammahat), force=1)

#refit the model
refit <- BayesPen.refit(y,W,fit.BayesPen)

#plot it
BayesPen.plot(refit)
```

BayesPen *Bayesian Penalized Credible Regions*

Description

Fits the Bayesian penalized credible regions method from the posterior mean and covariance of a Bayesian model. The function performs variable selection (Bondell and Reich 2012) and confounder selection (Wilson and Reich 2014+). The default is variable selection and confounder selection is only performed if confounder weights are provided.

Usage

```
BayesPen(beta, beta_cov, joint, force = NULL, confounder.weights, max.steps = NULL)
```

Arguments

`beta` p-vector of posterior means from the fitting of the full regression model. For confounder selection this is the outcome model.

`beta_cov` Posterior covariance matrix corresponding to beta.

`joint` Indicator if joint credible regions approach should be used. If joint=FALSE the marginal approach of Bondell and Reich (2012) will be used. The marginal approach is only available for variable selection, not confounder selection.

`force` A vector of columns corresponding to variables that are forced to be included in the model. For example, this may include an intercept. For confounder selection the exposure(s) of interest should be forced into the model.

`confounder.weights`
The posterior mean from the exposure model for confounder selection. For a single exposure this is a p-vector with exposure model regression coefficients in the same order as in beta. For multiple exposures this is a matrix with p rows and a column for each exposure. The locations corresponding to exposure(s) in beta can be set to any numeric value; they are not used. For variable selection this is omitted.

`max.steps` Maximum number of steps to be performed in the LARS algorithm (Hastie and Efron 2013).

Value

`joint.path` A complete solution path for the joint credible regions approach. Each row is a model in the solution path with a 1 indicating a variable is included and a 0 indicating it is not included.

`marginal.path` A complete solution path for the marginal credible regions approach. The p-vector denotes the step at which each covariate is included in the model.

`order.path` The action returned from lars that shows when each covariate is added to the model.

`order.marg` The the covariate added at each step.

`joint` Returns a vector indicating which variables are forced into the model.

`force` Returns the logical joint.

Author(s)

Ander Wilson, Howard D. Bondell, and Brian J. Reich

References

Bondell, H. D. and Reich, B. J. (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. *J. Am. Statist. Assoc.* 107, 1610-1624.

Trevor Hastie and Brad Efron (2013). lars: Least Angle Regression, Lasso and Forward Stagewise. R package version 1.2. http://CRAN.R-project.org/package=lars

Wilson, A. and Reich, B.J. (2014+). Confounder selection via penalized credible regions.

See Also

BayesPen.lm, BayesPen.lm.confounders

Examples

```
######################
#Variable Selection
set.seed(1234)
dat <- SimExample(500,model="BR1")
X <- dat$X
y <- dat$y

#fit the full model assuming flat priors on beta
fit1 <- lm(y~X-1)
betahat <- coef(fit1)
cov <- vcov(fit1)

#find solution path
fit.BayesPen <- BayesPen(beta=betahat, beta_cov=cov)

#refit the model
refit <- BayesPen.refit(y,X,fit.BayesPen)

#plot it
BayesPen.plot(refit)


######################
#Confounder Selection
set.seed(1234)
dat <- SimExample(500,model="WPD2")
X <- dat$X
U <- dat$U
W <- cbind(X,U)
y <- dat$y

#fit the full outcome model assuming flat priors on beta
fit1 <- lm(y~W-1)
betahat <- coef(fit1)
cov <- vcov(fit1)

#fit the full exposure model assuming flat priors on beta
```

```
fit2 <- lm(X~U-1)
gammahat <- coef(fit2)

#find solution path
fit.BayesPen <- BayesPen(beta=betahat, beta_cov=cov, confounder.weights=c(0,gammahat), force=1)

#refit the model
refit <- BayesPen.refit(y,W,fit.BayesPen)

#plot it
BayesPen.plot(refit)
```

`BayesPen.lm` *Variable Selection via Penalized Credible Regions for Linear Models*

Description

Fits a linear model using code based on BLR (Campos and Rodriguez 2012) and performs variable selection via penalized credible regions (Bondell and Reich 2012).

Usage

```
BayesPen.lm(y, x, prior, nIter, burnIn, thin, update, joint, force = NULL,
    max.steps = NULL, max.refit, saveAt = "")
```

Arguments

`y`	A n-vector of responses.
`x`	A n x p design matrix. If an intercept is desired then a column of ones should be included in the design matrix.
`prior`	A list containing the priors for the regression coefficients and the error variance. Each object in the list is also a list. The list prior$varE contains the degrees of freedom (prior$varE$df) and scale (prior$varE$S) for the inverse-X^2 prior on the error variance. The list prior$varBR contains the degrees of freedom (prior$varBR$df) and scale (prior$varBR$S) for the inverse-X^2 prior on the variance of the regression coefficients. In both cases the parameterization used has prior mean variance S/(df-2). See the documentation for BLR for additional details.
`nIter`	The number of MCMC iterations (integer).
`burnIn`	The number of MCMC iterations to be discarded as burnin (integer).
`thin`	Thinning number for the MCMC chain (integer)
`update`	Integer specifying how often to print an update on the progress of the MCMC.
`joint`	Indicator if joint credible regions approach should be used. If joint=FALSE the marginal approach of Bondell and Reich (2012) will be used.
`force`	An optional vector indexing which covariates variables should be forced into the model, for example an intercept may be forced into the model.

- `max.steps` Maximum number of steps to be performed in the LARS algorithm (Hastie and Efron 2013).
- `max.refit` The maximum number of models to be refit.
- `saveAt` An optional string that is a pre-fix for the filenames for the files saved while the program runs.

Value

- `joint.path` A complete solution path for the joint credible regions approach. Each row is a model in the solution path with a 1 indicating a variable is included and a 0 indicating it is not included.
- `marginal.path` A complete solution path for the marginal credible regions approach. The p-vector denotes the step at which each covariate is included in the model.
- `order.path` The action returned from lars that shows when each covariate is added to the model.
- `order.marg` The the covariate added at each step.
- `joint` Returns a vector indicating which variables are forced into the model.
- `force` Returns the logical joint.
- `coefs` A matrix of regression coefficients for each model in the solution path. The regression coefficients for parameters omitted from a model are set to 0.
- `SSE` SSE of each refitted model.
- `dev` Deviance of each refitted model.
- `df` Error degrees of freedom from each refitted model.
- `lm` Full fitting of the model.

Author(s)

Ander Wilson, Howard D. Bondell, and Brian J. Reich

References

Bondell, H. D. and Reich, B. J. (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. *J. Am. Statist. Assoc.* 107, 1610-1624.

Gustavo de los Campos and Paulino Perez Rodriguez, (2012). BLR: Bayesian Linear Regression. R package version 1.3. http://CRAN.R-project.org/package=BLR

Examples

```
set.seed(1234)
dat <- SimExample(500,model="BR1")
fit.BRnew <- BayesPen.lm(y=dat$y,x=dat$X)
BayesPen.plot(fit.BRnew)
```

`BayesPen.lm.confounders` *Confounder Selection via Penalized Credible Regions for Linear Models*

Description

Fits a linear outcome and exposure model using code based on BLR (Campos and Rodriguez 2012) and performs confounder selection via penalized credible regions (Wilson and Reich 2014+).

Usage

```
BayesPen.lm.confounders(y, x, u, prior, nIter, burnIn, thin, update, force,
       max.steps = NULL, max.refit, saveAt = "", include.me = FALSE,
       z.score = FALSE)
```

Arguments

`y`	A n-vector of responses.
`x`	For single exposures this is a n-vector of exposures. For multiple exposures this is a design matrix of the exposures with n rows and one column for each exposure.
`u`	A design matrix with potential confounders and other covariates. This should include a column of ones if an intercept is desired.
`prior`	A list containing the priors for the regression coefficients and the error variance. Each object in the list is also a list. The list prior$varE contains the degrees of freedom (prior$varE$df) and scale (prior$varE$S) for the inverse-X^2 prior on the error variance. The list prior$varBR contains the degrees of freedom (prior$varBR$df) and scale (prior$varBR$S) for the inverse-X^2 prior on the variance of the regression coefficients. In both cases the parameterization used has prior mean variance S/(df-2). See the documentation for BLR for additional details. Only one prior can be specified and will be used for both the outcome and exposure models.
`nIter`	The number of MCMC iterations (integer).
`burnIn`	The number of MCMC iterations to be discarded as burnin (integer).
`thin`	Thinning number for the MCMC chain (integer)
`update`	Integer specifying how often to print an update on the progress of the MCMC.
`force`	An optional vector indexing which confounding variables should be forced into the model, for example an intercept may be forced into the model. This vector indexes the columns of u. All exposures are forced into the model automatically.
`max.steps`	Maximum number of steps to be performed in the LARS algorithm (Hastie and Efron 2013).
`max.refit`	The maximum number of models to be refit.

`saveAt` An optional string that is a pre-fix for the filenames for the files saved while the program runs.

`include.me` Indicator for the multiple exposure case indicating if the exposure models should include the other exposures as covariates.

`z.score` Indicator for using z-scores in the penalty instead of regression coefficients. If the error variance for the exposure and outcome models or the scale of the regression coefficients is very different then using the z-scores make the confounder and exposure components of the weights more comparable.

Value

`joint.path` A complete solution path for the joint credible regions approach. Each row is a model in the solution path with a 1 indicating a variable is included and a 0 indicating it is not included.

`marginal.path` Not used for confounder selection.

`order.path` The action returned from lars that shows when each covariate is added to the model.

`order.marg` Not used for confounder selection.

`joint` Always TRUE for confounder selection.

`force` Vector of variables forced into the model including the exposures.

`coefs` A matrix of regression coefficients for each model in the solution path. The regression coefficients for parameters omitted from a model are set to 0.

`SSE` SSE of each refitted model.

`dev` Deviance of each refitted model.

`df` Error degrees of freedom from each refitted model.

`lm` Full fitting of the outcome model.

`confounder.weights`
Confounder weights used in BayesPen to fit the model.

Author(s)

Ander Wilson, Howard D. Bondell, and Brian J. Reich

References

Gustavo de los Campos and Paulino Perez Rodriguez, (2012). BLR: Bayesian Linear Regression. R package version 1.3. http://CRAN.R-project.org/package=BLR

Wilson, A. and Reich, B.J. (2014+). Confounder selection via penalized credible regions.

Examples

```
set.seed(1234)
dat <- SimExample(500,model="WPD2")
fit.BRnew <- BayesPen.lm.confounders(y=dat$y,x=dat$X, u=dat$U)
BayesPen.plot(fit.BRnew)
```

BayesPen.plot — *Plot Bayesian Penalized Credible Region Solution Path.*

Description

Plots the solution path for Bayesian penalized credible regions from a an object returned by BayesPen.refit.

Usage

```
BayesPen.plot(refit, ...)
```

Arguments

`refit`	Object returned from BayesPen.refit.
`...`	Additional graphics parameters from matplot.

Author(s)

Ander Wilson, Howard D. Bondell, and Brian J. Reich

References

Bondell, H. D. and Reich, B. J. (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. *J. Am. Statist. Assoc*. 107, 1610-1624.

Wilson, A. and Reich, B.J. (2014+). Confounder selection via penalized credible regions.

See Also

`BayesPen`,`BayesPen.refit`

Examples

```
set.seed(1234)
dat <- SimExample(500,model="WPD2")
fit.BRnew <- BayesPen.lm.confounders(y=dat$y,x=dat$X, u=dat$U)
BayesPen.plot(fit.BRnew,
  lty=1, #change line type
  lwd=c(10,rep(2,nrow(fit.BRnew$coef)-1)), #change width
  col=c("black","grey",rep("blue",7),rep("red",7),
  rep("grey",nrow(fit.BRnew$coef)-15)), #change the colors
   ylim=c(-.1,.4) #set limits
  )
legend("topright", lty=1, lwd=c(10,2,2,2), col=c("black","blue","red","grey"),
   legend=c("Exposure","Confounder","Covariate","Noise"))
```

BayesPen.refit *Bayesian Penalized Credible Regions Solution Path Refit*

Description

Refits the solution path given by BayesPen.

Usage

```
BayesPen.refit(y, x, fit, joint, max.refit, ...)
```

Arguments

`y`	A n-vector of responses. If fit is a list from BayesPen.lm or BayesPen.lm.confounders then y is not required.
`x`	A n x p design matrix that includes all potential covariates. In the confounder selection case this includes the exposures and confounders, i.e. cbind(x,u). If fit is a list from BayesPen.lm or BayesPen.lm.confounders then x is not required.
`fit`	A list returned from BayesPen, BayesPen.lm, or BayesPen.lm.confounders.
`joint`	For variable selection this indicates if the joint or marginal solution path should be used. Joint must be TRUE for confounder selection.
`max.refit`	The maximum number of models to be refit.
`...`	These are additional terms passed to glm to refit a glm other than linear. The default is linear.

Details

This refits each model in the solution path with the frequentist model using the glm function.

Value

`coefs`	A matrix of regression coefficients for each model in the solution path. The regression coefficients for parameters omitted from a model are set to 0.
`SSE`	SSE of each refitted model.
`dev`	Deviance of each refitted model.
`df`	Error degrees of freedom from each refitted model.
`joint`	Returns the logical joint.

Author(s)

Ander Wilson, Howard D. Bondell, and Brian J. Reich

References

Bondell, H. D. and Reich, B. J. (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. *J. Am. Statist. Assoc.* 107, 1610-1624.

Wilson, A. and Reich, B.J. (2014+). Confounder selection via penalized credible regions.

See Also

BayesPen

Examples

```
######################
#Variable Selection
set.seed(1234)
dat <- SimExample(500,model="BR1")
X <- dat$X
y <- dat$y

#fit the full model assuming flat priors on beta
betahat <- solve(t(X)%*%X) %*% t(X) %*% y
cov <- solve(t(X)%*%X) * sum((X%*%betahat-y)^2)/(length(y)-length(betahat))

#find solution path
fit.BayesPen <- BayesPen(beta=betahat, beta_cov=cov)

#refit the model
refit <- BayesPen.refit(y,X,fit.BayesPen)

#plot it
BayesPen.plot(refit)
```

SimExample *Simulate Example Data for BayesPen*

Description

Simulates example data used in Bondell and Reich (2012), Wang et. al (2012), and Wilson and Reich (2014+).

Usage

```
SimExample(n = 100, p, model, rho)
```

Arguments

`n`	Number of observations.
`p`	The total number of covariates (including the exposure of interest for WPD2).
`model`	What model to simulate data from. WPD2 is design 2 from Wang et. al (2012). BR1 and BR2 are designs 1 and 2 from Bondell and Reich (2012).
`rho`	This specifies the correlation between covariates in WPD2 and BR2.

Value

`y`	n vector of responses.
`X`	n vector of exposures for WPD and the n x p design matrix for BR1 and BR2.
`U`	n x p matrix of potential confounders for WPD2.This is missing for BR1 and BR2.
`p`	Total number of potential confounders
`beta`	The regression coeficients. For WPD2 the first beta corresponds to X.
`rho`	Returns rho.
`model`	Returns model.

Author(s)

Ander Wilson, Howard D. Bondell, and Brian J. Reich

References

Bondell, H. D. and Reich, B. J. (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. *J. Am. Statist. Assoc.* 107, 1610-1624.

Wang, C., Parmigiani, G., and Dominici, F. (2012). Bayesian effect estimation accounting for adjustment uncertainty. *Biometrics* 68, 661-671.

Examples

```
set.seed(1234)
dat <- SimExample(500,model="BR1")
lm.fit <- lm(dat$y~dat$X)
```

Index