Package 'CBPS'

February 15, 2013

Version 0.3

Date 2013-01-18

Title R Package for Covariate Balancing Propensity Score

Author Marc Ratkovic <ratkovic@princeton.edu>, Kosuke Imai
<kimai@princeton.edu>, Christian Fong <cjfong@princeton.edu>

Maintainer Marc Ratkovic <ratkovic@princeton.edu>

Depends R (>= 2.14), MASS, arm, MatchIt

Imports arm

Description CBPS is an R package that implements the covariate
balancing propensity score proposed by Imai and Ratkovic
(2012). The propensity score is estimated such that it
maximizes the resulting covariate balance as well as the prediction of treatment assign-
ment. The method, therefore,avoids an iteration between model fitting and balance checking.
Currently, the package can only deal with binary treatment in
the cross section setting but in the future it will be extended
to other settings including non-binary treatments and causal inference with longitudinal data.

LazyLoad yes

LazyData yes

License GPL (>= 2)

Repository CRAN

Date/Publication 2013-01-31 17:07:43

NeedsCompilation no

R topics documented:

CBPS   *Covariate Balancing Propensity Score (CBPS) Estimation*

Description

CBPS and CBPS.fit estimate propensity score such that both covariate balance and prediction of treatment assignment are maximized. The method, therefore, avoids an iterative process between model fitting and balance checking and implements both simultaneously.

Usage

```
CBPS(formula, data, na.action, ATT, method = "over", ...)
CBPS.fit(treat, X, ATT, X.bal = X, method, ...)
```

Arguments

formula	An object of class formula (or one that can be coerced to that class): a symbolic description of the model to be fitted.
data	an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which CBPS is called.
na.action	A function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset.
ATT	Set to TRUE to find the average treatment effect on the treated, and FALSE to find the average treatment effect. Default is TRUE.
treat	A vector of treatment assignments.
X	A covariate matrix.
X.bal	A matrix to be balanced. Currently not accepted.
method	"over" for fitting an over-identified model that combines the propensity score and covariate balancing conditions; "exact" for fitting a model that only contains the covariate balancing conditions.
...	Other parameters to be passed through to optim()

Details

Fits covariate balancing propensity scores.

Value

coefficients	A named vector of coefficients
residuals	The working residuals from the final iteration of CBPS
fitted.values	The fitted mean values, obtained by transforming the linear predictors by the inverse of the binomial function
rank	The numeric rank of the fitted model
family	The family object used.
deviance	Minus twice the log-likelihood of the CBPS fit. Will be lower than the maximum likelihood deviance.
weights	The CBPS balancing weights calculated from the propensity scores.
y	The treatment vector used
x	The covariate matrix
model	The model frame
converged	Convergence value. Returned from the call to optim().
call	The matched call
formula	The formula supplied.
data	The data argument.
J	The J-statistic at convergence.
df	The degrees of freedom.
var	The covariance matrix, evaluated numerically from optim().

Author(s)

Marc Ratkovic, Kosuke Imai, and Christian Fong; The CBPS function is based on the code for version 2.15.0 of the glm function implemented in the stats package, originally written by Simon Davies. This documenation is likewise modeled on the documentation for glm and borrows its language where the arguments and values are the same.

References

Imai, Kosuke and Marc Ratkovic. "Covariate Balancing Propensity Score." Unpublished Manuscript, Princeton University. <http://imai.princeton.edu/research/CBPS.html>

See Also

summary.CBPS

Examples

```
## Not run:
###
### Example: propensity score matching
###

##Load the LaLonde data
data(LaLonde)
## Estimate CBPS via logistic regression
fit <- CBPS(treat ~ age + educ + re75 + re74 + I(re75==0) + I(re74==0), data = LaLonde, ATT = TRUE)
summary(fit)
## matching via MatchIt: one to one nearest neighbor with replacement
library(MatchIt)
m.out <- matchit(treat ~ 1, distance = fitted(fit), method = "nearest", data = LaLonde, replace = TRUE)


### Example: propensity score weighting
###
## Simulation from Kang and Shafer (2007).
set.seed(123456)
n <- 1000
X <- mvrnorm(n, mu = rep(0, 4), Sigma = diag(4))
prop <- 1 / (1 + exp(X[,1] - 0.5 * X[,2] + 0.25*X[,3] + 0.1 * X[,4]))
treat <- rbinom(n, 1, prop)
y <- 210 + 27.4*X[,1] + 13.7*X[,2] + 13.7*X[,3] + 13.7*X[,4] + rnorm(n)

##Estimate CBPS with a misspecificied model
X.mis <- cbind(exp(X[,1]/2), X[,2]*(1+exp(X[,1]))^(-1)+10, (X[,1]*X[,3]/25+.6)^3, (X[,2]+X[,4]+20)^2)
fit1 <- CBPS(treat ~ X.mis, ATT = TRUE)

## Horwitz-Thompson estimate
mean(treat*y/fit1$fitted.values)
## Inverse probability weighting
sum(treat*y/fit1$fitted.values)/sum(treat/fit1$fitted.values)

## End(Not run)
```

DR   *Doubly Robust Estimator*

Description

DR calculates the doubly-robust estimator as as described in Lunceford and Davidian (2004) and originally detailed in Robins, Rotnitzky, and Zhao (1994).

Usage

```
DR(formula,model,data,treat,pscore)
```

Arguments

formula	An object of class formula (or one that can be coerced to that class): a symbolic description of the model to be fitted.
model	Set to "lm" to use a linear model, set to "glm" for the generalized linear model.
data	An optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which CBPS is called.
treat	A vector of treatment assignments, where 1 indicates assignment to the treatment group and 0 indicates assignment to the control group.
pscore	A vector of propensity scores.

Details

Fits covariate balancing propensity scores.

Value

Returns a vector with the point estimate of the doubly-robust estimator as well as its standard error. For use with CBPS, run only with the ATE.

Author(s)

Marc Ratkovic, Kosuke Imai, and Christian Fong.

References

Lunceford, Jared and Marie Davidian. "Stratification and Weighting Via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study." Statistics in Medicine 23, 15 October 2004.

See Also

IPW

Examples

```
## Not run:
###
### Example: Doubly robust estimator
###

##Load the LaLonde data
data(LaLonde)
## Estimate CBPS via logistic regression for ATE.  Run only with ATE.
fit <- CBPS(treat ~ age + educ + re75 + re74 + I(re75==0) + I(re74==0), data = LaLonde, ATT = FALSE)
## Find doubly robust estimator under GLM
doubly.robust <- DR(re78 ~ age + educ + re75 + re74 + I(re75==0) + I(re74==0), model="glm", data = LaLonde, tr
```

```
## End(Not run)
```

IPW   *Inverse Probability Weighting Estimators*

Description

`IPW` calculates the inverse probability weighting estimators from Lunceford and Davidian (2004).

Usage

```
IPW(outcome, treat, data, pscore, k)
```

Arguments

outcome	A vector of the outcome variable.
treat	A vector of treatment assignments, where 1 indicates assignment to the treatment group and 0 indicates assignment to the control group.
data	an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which CBPS is called.
pscore	A vector of propensity scores.
k	The number of covariates considered in the generation of the propensity scores.

Details

Calculates the inverse probability weighting estimators from Lunceford and Davidian (2004). For use with CBPS, run only with ATE.

Value

Returns a data frame whose columns are the point estimates and standard errors for these inverse probability weighting estimators, and whose rows correspond to IPW1, IPW2, and IPW3.

Author(s)

Marc Ratkovic, Kosuke Imai, and Christian Fong.

References

Lunceford, Jared and Marie Davidian. "Stratification and Weighting Via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study." Statistics in Medicine 23, 15 October 2004.

See Also

DR

Examples

```
## Not run:
###
### Example: propensity score matching
###

##Load the LaLonde data
data(LaLonde)
## Estimate CBPS via logistic regression for ATE (Run with ATE only)
fit <- CBPS(treat ~ age + educ + re75 + re74 + I(re75==0) + I(re74==0), data = LaLonde, ATT = FALSE)
##Calculate the IPW estimators
inverse.prob.wt <- IPW(outcome = re78, treat = treat, data = LaLonde, pscore = fitted(fit), k = length(coef(fi

## End(Not run)
```

LaLonde *LaLonde Data for Covariate Balancing Propensity Score*

Description

This data set gives the outcomes a well as treatment assignments and covariates for the econometric evaluation of training programs in LaLonde (1986).

Usage

```
LaLonde
```

Format

A data frame consisting of 5 columns (including a treatment assignment vector) and 2787 observations.

Source

Data from the National Supported Work Study. A benchmark matching dataset. Columns consist of an indicator for whether the observed unit was in the experimental subset; an indicator for whether the individual received the treatment; age in years; schooling in years; indicators for black and Hispanic; an indicator for marriage status, one of married; an indicator for no high school degree; and reported earnings in 1974, 1975, and 1978. 1974 and 1975 earnings are pre-treatment. 1978 earnings is taken as the outcome variable.

References

LaLonde, R.J. (1986). Evaluating the econometric evaulations of training programs with experimental data. American Economic Review 76, 4, 604-620.

---

summary.CBPS   *Summarizing Covariate Balancing Propensity Score Estimation*

Description

Prints a summary of a fitted CBPS object.

Usage

```
## S3 method for class 'CBPS'
summary(object,...)
```

Arguments

object	an object of class "CBPSlogit", usually, a result of a call to CBPS.
...	Additional arguments to be passed to summary.

Details

Prints a summmary of a CBPS object, in a format similar to glm. The variance matrix is calculated from the numerical Hessian at convergence of CBPS.

Value

call	The matched call.
deviance.residuals	
	The five number summary and the mean of the deviance residuals.
coefficients	A table including the estimate for the each coefficient and the standard error, z-value, and two-sided p-value for these estimates.
J	Hansen's J-Statistic for the fitted model.
Log-Likelihood	The log-likelihood of the fitted model.

Author(s)

Marc Ratkovic

See Also

CBPS, summary

---

vcov.CBPS   *Calculate Variance-Covariance Matrix for a Fitted CBPS Object*

Description

vcov.CBPS Returns the variance-covariance matrix of the main parameters of a fitted CBPS object.

Usage

```
## S3 method for class 'CBPS'
vcov(object,...)
```

Arguments

object	An object of class formula (or one that can be coerced to that class): a symbolic description of the model to be fitted.
...	Additional arguments to be passed to vcov.CBPS

Details

This is the CBPS implementation of the generic function vcov().

Value

A matrix of the estimated covariances between the parameter estimates in the linear or non-linear predictor of the model.

Author(s)

Marc Ratkovic, Kosuke Imai, and Christian Fong.

References

This documentation is modeled on the documentation of the generic vcov.

See Also

vcov

Examples

```
###
### Example: Variance-Covariance Matrix
###

##Load the LaLonde data
data(LaLonde)
## Estimate CBPS via logistic regression
fit <- CBPS(treat ~ age + educ + re75 + re74 + I(re75==0) + I(re74==0), data = LaLonde, ATT = TRUE)
```

```
## Get the variance-covariance matrix.
vcov(fit)
```

Index