

Manual of the *BayesPSA* Package

Jianshen Chen¹, Fabrizio Sanchez² & David Kaplan²

List of methods

- `bpsa()`
 - `bpsa.strat()`
 - `bpsa.opt.match()`
 - `cov.balance()`
 - `cov.bal.plot()`
-

`bpsa()`

Description

The first step in performing a Bayesian propensity score analysis. Runs a Bayesian logistic regression on the treatment factor, either via MCMC or Bayesian Model Averaging (BMA).

Usage

```
bpsa(x, response, treatment, treatment.success.level = 1, vars = names(x), exclude = NULL, prior.b0 = 0, prior.B0 = 0, mcmc.burnin = 1250, mcmc.iter = 1e+05, mcmc.thin = 100, mcmc.logfun = TRUE, mcmc.tune = 0.25, method = c("MCMC", "BMA"))
```

Arguments

x a data frame containing the response, treatment and covariates for the propensity score estimation.

response the name of the dependent variable one is interested in for the causal model. Can be either a column name (entered as a character string) or a column number (entered as an integer).

¹Educational Testing Service, jchen006@ets.org; ² University of Wisconsin–Madison; The research reported in this paper was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D110001 to The University of Wisconsin - Madison. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the U.S. Department of Education.

treatment the treatment variable. It is expected that ‘treatment’ will be a two-level factor. Can be either a column name (entered as a character string) or a column number (entered as an integer).

treatment.success.level which level (1 or 2) of the ‘treatment’ factor is considered to be the ‘success’ or ‘treatment’ level. Default is set to 1.

vars names of the variables to be used to calculate the propensity score. Default is set to be all the columns in x (other than ‘response’ and ‘treatment’). Can be either a vector of column names or a vector of column numbers.

exclude an optional vector of column names or vector of column indices to exclude from the logistic regression.

prior.b0 prior mean of beta under a multivariate normal prior. This can either be a scalar or a column vector with dimension equal to the number of betas. If this takes a scalar value, then that value will serve as the prior mean for all of the betas.

prior.B0 prior precision of beta under a multivariate normal prior. This can either be a scalar or a square matrix with dimensions equal to the number of betas. If this takes a scalar value, then that value times an identity matrix serves as the prior precision of beta. Default value of 0 is equivalent to an improper uniform prior for beta.

mcmc.burnin if method == ‘MCMC’, the number of iterations to throw away before sampling from the chain. Default is set to 5,000. Ignored if method == ‘BMA’.

mcmc.iter if method == ‘MCMC’, the number of iterations in the MCMC chain. Default is set to 400,000. Ignored if method == ‘BMA’.

mcmc.thin if method == ‘MCMC’, the thinning interval of the MCMC chain. Default is set to 400, which yields 1000 posterior draws in 400,000 iterations. Ignored if method == ‘BMA’.

mcmc.logfun default to TRUE. See *?MCMClogit* for more details.

mcmc.tune tuning parameter for the MCMC sampling. Default to 0.25. See *?MCMClogit* for more details.

Returned Values

p.score a matrix containing the estimated propensity scores.

mcmc.model if method == ‘MCMC’, the model created by MCMClogit. Otherwise NULL.

mcmc.model.matrix if method == ‘MCMC’, the model created by MCMClogit. Otherwise NULL.

- x.data** the original data frame used to in the `bpsa()` call.
- x.vars** a vector containing the names of the variables used as co-variates in creating the propensity score.
- response** the name of the response variable.
- treatment** the name of the treatment variable.

References

- R Development Core Team. (2011). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org> (ISBN3-900051-07-0)
- Martin, A. D., Quinn, K. M., & Park, J. H. (2010). Markov chain Monte Carlo (MCMC) package. Retrieved from <http://mcmcpack.wustl.edu/>
- Raftery, A. E., Hoeting, J., Volinsky, C., Painter, I., & Yeung, K. Y. (2009). Bayesianmodel averaging (BMA), version 3.12. Retrieved from <http://www2.research.att.com/volinsky/bma.html>
- Kaplan, D., & Chen, J. (2012). A two-step Bayesian approach for propensity score analysis: Simulations and case study. *Psychometrika*, 77, 581 – 609.

Examples

See example in the help page for ‘bpsa’ using ‘*?bpsa*’.

`bpsa.strat()`

Description

Performs a stratified matching using a propensity score created by calling `bpsa()`.

Usage

```
bpsa.strat(x, prior.b0 = 0, prior.B0 = 0, mcmc.burnin = 1000, mcmc.iter = 1000, mcmc.thin = 1)
```

Arguments

x a `bpsa` list created by calling the function `'bpsa()'`.

prior.b0 prior mean of beta under a multivariate normal prior. This can either be a scalar or a column vector with dimension equal to the number of betas. If this takes a scalar value, then that value will serve as the prior mean for all of the betas.

prior.B0 prior precision of beta under a multivariate normal prior. This can either be a scalar or a square matrix with dimensions equal to the number of betas. If this takes a scalar value, then that value times an identity matrix serves as the prior precision of beta. Default value of 0 is equivalent to an improper uniform prior for beta.

mcmc.burnin the number of iterations to throw away before sampling from the MCMC chain in the Bayesian outcome model. Default is set to 100.

mcmc.iter the number of iterations in the MCMC chain of the Bayesian outcome model. Default is set to 1,000.

mcmc.thin the thinning interval of the MCMC chain in the Bayesian outcome model. Default is set to 1.

Returned Values

MATCHING.METHOD a character string `'strat'`, indicating stratification method is utilized.

EFFECT the posterior mean and standard deviation of causal(treatment) effect.

INTERVAL the 95% posterior probability interval.

References

- Kaplan, D., & Chen, J. (2012). A two-step Bayesian approach for propensity score analysis: Simulations and case study. *Psychometrika*, 77, 581 – 609.
- Martin, A. D., Quinn, K. M., & Park, J. H. (2010). Markov chain Monte Carlo (MCMC) package. Retrieved from <http://mcmcpack.wustl.edu/>

Examples

See example in the help page for `'bpsa.strat'` using `'?bpsa.strat'`.

`bpsa.opt.match()`

Description

Performs an optimal full matching using a propensity score created by calling `bpsa()`. It can be quite time-consuming when sample size is large.

Usage

```
bpsa.opt.match(x, prior.b0 = 0, prior.B0 = 0, mcmc.burnin = 1000, mcmc.iter = 10000, mcmc.thin = 10)
```

Arguments

x a `bpsa` list created by calling the function `'bpsa()'`.

prior.b0 prior mean of beta under a multivariate normal prior. This can either be a scalar or a column vector with dimension equal to the number of betas. If this takes a scalar value, then that value will serve as the prior mean for all of the betas.

prior.B0 prior precision of beta under a multivariate normal prior. This can either be a scalar or a square matrix with dimensions equal to the number of betas. If this takes a scalar value, then that value times an identity matrix serves as the prior precision of beta. Default value of 0 is equivalent to an improper uniform prior for beta.

mcmc.burnin the number of iterations to throw away before sampling from the MCMC chain in the Bayesian outcome model. Default is set to 100.

mcmc.iter the number of iterations in the MCMC chain of the Bayesian outcome model. Default is set to 1,000.

mcmc.thin the thinning interval of the MCMC chain in the Bayesian outcome model. Default is set to 1.

Returned Values

MATCHING.METHOD a character string `'optimal'`, indicating optimal matching method is utilized.

MATCHES.MATRIX a N by M matrix showing the labels of matched groups, where N refers to number of subjects and M refers to number of posterior draws of the propensity score. In the example, N=200 and M=1000.

EFFECT the posterior mean and standard deviation of causal(treatment) effect.

INTERVAL the 95% posterior probability interval.

References

- Hansen, B. B., Klopfer, S. O. (2006). Optimal full matching and related designs via network flow. *Journal of Computational and Graphical Statistics*, 15, 609 – 627.
- Kaplan, D., & Chen, J. (2012). A two-step Bayesian approach for propensity score analysis: Simulations and case study. *Psychometrika*, 77, 581 – 609.
- Martin, A. D., Quinn, K. M., & Park, J. H. (2010). Markov chain Monte Carlo (MCMC) package. Retrived form [http : //mcmcpack.wustl.edu/](http://mcmcpack.wustl.edu/)

Examples

See example in the help page for ‘bpsa.opt.match’ using ‘*?bpsa.opt.match*’.

cov.balance()

Description

Checks the balance of independent variables (covariates) used in obtaining the propensity score.

Usage

```
cov.balance(x.bpsa, x.data = NULL, x.match, check.vars = x.bpsavar.argsX.VARS,  
type.vars = NULL)
```

Arguments

x.bpsa the bpsa object used for covariate balance checking.

x.data user specified data to perform balance check upon. Default NULL is to use the dataset where the propensity score was created from.

x.match the type of matching that we want to check the covariate balance upon. Can be the output object of *bpsa.strat()* for Bayesian propensity score stratification or *bpsa.opt.match()* for Bayesian optimal matching.

check.vars vector of the names of which covariates in x.data we want to check.

type.vars vector of characters indicating whether variables in 'check.vars' are factors or not (TRUE, FALSE) respectively. Default NULL assumes all covariates are continuous variables. Must be the same length at check.vars.

Returned Values

METHOD a character string 'MCMC' or 'BMA', indicating which estimation method is utilized.

IMBALANCE a 7 by K matrix indicating covariate balance performance, including mean of each covariate in each of the two groups, mean difference of each covariate in two groups, covariate variance in each of the two groups, Cohen's d/standardized mean difference (proportion difference for a binary variable) and variance ratio. The number of columns, K, refers to the number of covariates being checked.

VAR.TYPE a character string indicating the covariate types, binary or continuous.

INIT.COHEN.IMBALANCE initial Cohen's d averaged over all the covariates.

INIT.VARRATIO.IMBALANCE initial variance ratio averaged over all the covariates.

COHEN.MCMC a M by K matrix showing Cohen's d for each covariate based on each posterior draw of the propensity score. M refers to the number of posterior draws of the propensity score and K indicates the number of covariates.

COHEN.MEANS Cohen's d averaged over all the posterior draws of the propensity score and all the covariates.

COHEN.QUANTILES 95% posterior probability interval of Cohen's d averaged over all the covariates.

MEAN.DIFF.MCMC a M by K matrix of mean difference for each covariate based on each posterior draw of the propensity score.

VAR.RATIO.MCMC a M by K matrix showing variance ratio for each covariate based on each posterior draw of the propensity score. For some matched groups with only one subject in the treatment or control group, the variances of the covariates will be 0 and in turn an infinite variance ratio will be obtained.

VAR.RATIO.MEANS Variance ratio averaged over all the posterior draws of the propensity score and all the covariates. By default, the extreme variance ratios (if any, e.g., infinite) are included in the mean calculation. Users can calculate the average variance ratio in other ways based on the variance ratio matrix by calling ‘VAR.RATIO.MCMC’.

VAR.RATIO.QUANTILE 95% posterior probability interval of variance ratio averaged over all the covariates.

References

- Chen, J., & Kaplan, D. (2014). Covariate balance in Bayesian propensity score approaches for observational studies. *Journal for Research on Educational Effectiveness*. *doi* : 10.1080/19345747.2014.911396

Examples

See examples for *bpsa.opt.match()*.

cov.bal.plot()

Description

Produce a plot for covariate balance checking using Cohen’s d and variance ratio of covariates in two different treatment conditions (e.g., treatment group and control group).

Usage

```
cov.bal.plot(bal.obj, show.var.names = FALSE, point.size = 3)
```

Arguments

bal.obj a covariance balance check object obtained by calling *cov.balance()*.

show.var.names set to TRUE to see variable names in plot.

point.size sets the size of the points in the *ggplot*.

Returned Values

Initial covariate balance plot and covariate balance plot after Bayesian propensity score adjustment in the pop-up plot windows.

References

- Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. Springer New York. [http : //had.co.nz/ggplot2/book](http://had.co.nz/ggplot2/book)
- Chen, J., & Kaplan, D. (2014). Covariate balance in Bayesian propensity score approaches for observational studies. *Journal for Research on Educational Effectiveness*. *doi* : 10.1080/19345747.2014.911396

Examples

See examples for *bpsa.opt.match()*.