Package 'pstest'

November 20, 2016

Type Package

Title Specification Tests for Parametric Propensity Score Models

Version 0.1.2.901

Author Pedro H. C. Sant'Anna <pedro.h.santanna@vanderbilt.edu>, Xiajoun Song

<sxj@gsm.pku.edu.cn>

Maintainer Pedro H. C. Sant'Anna <pedro.h.santanna@vanderbilt.edu>

Description The propensity score is one of the most widely used tools in studying the causal effect
of a treatment, intervention, or policy. Given that the propensity score is usually unknown,
it has to be estimated, implying that the reliability of many treatment effect estimators depends
on the correct specification of the (parametric) propensity score. This package provides
data-driven nonparametric diagnostic tools for detecting propensity score misspecification.

URL https://github.com/pedrohcgs/pstest

Depends R (>= 3.1)

Imports stats,
parallel,
harvestr

License GPL-2

LazyData TRUE

RoxygenNote 5.0.1.9000

Date 2016-11-19 19:35

R topics documented:

1

pstest-package	*pstest: An R Package for assessing the goodness-of-fit of parametric propensity score models.*

Description

The propensity score is one of the most widely used tools in studying the causal effect of a treatment, intervention, or policy. Given that the propensity score is usually unknown, it has to be estimated, implying that the reliability of many treatment effect estimators depends on the correct specification of the (parametric) propensity score. This package provides data-driven nonparametric diagnostic tools for detecting propensity score misspecification.

Details

This R package implements the class of specification test for the propensity score proposed in Sant'Anna and Song (2016), 'Specification Tests for the Propensity Score', available at Pedro H.C. Sant'Anna webpage, http://sites.google.com/site/pedrohcsantanna/.

In short, this package implements Kolmogorov-Smirnov and Cramer-von Mises type tests for parametric propensity score models with either logistic ('logit'), or normal ('probit') link function. Critical values are computed with the assistance of a multiplier bootstrap.

The tests are based on the integrated conditional moment approach, where the weight function used is based on an orthogonal projection onto the tangent space of nuisance parameters. As a result, the tests (a) enjoy improved power properties, (b) do not suffer from the 'curse of dimensionality' when the vector of covariates is of high-dimensionality, (c) are fully data-driven, (e) do not require tuning parameters such as bandwidths, and (e) are able to detect a broad class of local alternatives converging to the null at the parametric rate. These appealing features highlight that the tests can be of great use in practice.

It is worth stressing that this package implements in a unified manner a large class of specification tests, depending on the chosen weight function $w(q, u)$:

- 'ind' - the indicator weight function $w(q, u) = 1(q \leq u)$. This is the default.
- 'exp' - the exponential weight function $w(q, u) = exp(qu)$.
- 'logistic' - the logistic weight function $w(q, u) = 1/[1 + exp(-qu)]$.
- 'sin' - the sine weight function $w(q, u) = sin(qu)$.
- 'sincos' - the sine and cosine weight function $w(q, u) = sin(qu) + cos(qu)$.

Different weight functions $w(q, u)$ have different power properties, and therefore, being able to choose different $w(q, u)$ gives us flexibility to direct power in desired directions.

References

Sant'Anna, Pedro H. C, and Song, Xiaojun (2016), *Specification Tests for the Propensity Score*, available at http://sites.google.com/site/pedrohcsantanna/.

---

pstest *pstest: Tests for the Propensity Score*

Description

pstest computes Kolmogorov-Smirnov and Cramer-von Mises type tests for the null hypothesis that a parametric model for the propensity score is is correctly specified. For details of the testing procedure, see Sant'Anna and Song (2016),'Specification Tests for the Propensity Score'.

Usage

```
pstest(d, pscore, xpscore, model = c("logit", "probit"), w = c("ind", "exp",
  "logistic", "sin", "sincos"), dist = c("Mammen", "Rademacher"),
  nboot = 1000, cores = 1, chunk = 1000)
```

Arguments

d	a vector containing the binary treatment indicator.
pscore	a vector containing the estimated propensity scores.
xpscore	a matrix (or data frame) containing the covariates (and their transformations) included in the propensity score estimation. It should also include the constant term.
model	a description of the functional form (link function) used to estimated propensity score. The alternatives are: 'logit' (default), and 'probit'.
w	a description of which weight function the projection is based on. The alternatives are 'ind' (default), which set $w(q, u) = 1(q <= u)$, 'exp', which set $w(q, u) = exp(qu)$, 'logistic', which set $w(q, u) = 1/[1 + exp(-qu)]$, 'sin', which set $w(q, u) = sin(qu)$, and 'sincos', which set $w(q, u) = sin(qu) + cos(qu)$.
dist	a description of which distribution to use during the bootstrap. The alternatives are 'Mammen' (default), and 'Rademacher'.
nboot	number of bootstrap replicates to perform. Default is 1,000.
cores	number of cores to use during the bootstrap. Default is 1. If cores is greater than 1, the bootstrap is conducted using parLapply, instead of lapply type call.
chunk	a value that determine the size of each 'tile'. Such argument is used to split the original data into chunks, saving memory. Default value is 1,000. If the *pstest* function throw a memory error, you should choose a smaller value for *chunk*.

Value

a list containing the Kolmogorov-Smirnov and Cramer-von Mises test statistics for the null hypothesis of correctly specified propensity score model (kstest and cvmtest, respectively), and their associate bootstrapped p-values, pvks and pvcvm, respectively. All inputs are also returned.

References

Sant'Anna, Pedro H. C, and Song, Xiaojun (2016), *Specification Tests for the Propensity Score*, available at http://sites.google.com/site/pedrohcsantanna/.

Examples

```
# Example based on simulation data
# Simulate vector of covariates
set.seed(1234)
x1 <- runif(100)
x2 <- rt(100, 5)
x3 <- rpois(100, 3)
# generate treatment status score based on Probit Specification
treat <- (x1 + x2 + x3 >= rnorm(100, 4, 5))
# estimate correctly specified propensity score based on Probit
pscore <- stats::glm(treat ~ x1 + x2 + x3, family = binomial(link = "probit"),
             x = TRUE)
# Test the correct specification of estimated propensity score, using
# the weight function 'ind', and bootstrap based on 'Mammen'.
pstest(d = pscore$y, pscore = pscore$fit, xpscore = pscore$x,
       model = "probit", w = "ind", dist = "Mammen")
# Alternatively, one can use the 'sin' weight function
pstest(d = pscore$y, pscore = pscore$fit, xpscore = pscore$x,
       model = "probit", w = "sin", dist = "Mammen")
```

Index