

Package ‘EstACEwErrorinX’

November 24, 2014

Type Package

Title Average Causal Effect (ACE) Estimation with Covariate Measurement Error

Version 1.1

Date 2014-11-15

Author Andrew M. Raim, Xiaoyu Dong, Yi Huang

Corresponding Author Yi Huang <yihuang@umbc.edu>

Maintainer Andrew M. Raim <araim1@umbc.edu>

Description

License TBD

Depends MASS, MatchIt

R topics documented:

| | |
|------------------------------------|----|
| EstACEwErrorinX-package | 2 |
| breastpump | 3 |
| cov.louis | 4 |
| d.group | 5 |
| d.mixture | 6 |
| d.mixture.x | 6 |
| EM | 7 |
| FormHessian.Q3 | 9 |
| FormScore.Q3 | 9 |
| naive.ace | 10 |
| new.ace | 10 |
| NewtonRaphson.Q3 | 11 |
| NewtonRaphson.Q3.OneStep | 12 |
| Pi.matrix | 12 |
| Q3 | 13 |
| r.mixture | 13 |

| | |
|--------------|-----------|
| Index | 15 |
|--------------|-----------|

EstACEwErrorinX-package

Average Causal Effect (ACE) Estimation with Covariate Measurement Error

Description

Average causal effect (ACE) estimation allowing covariate measurement error using newly developed latent propensity scoring approach in a finite mixture modeling framework.

Details

Although an analysis based on a carefully conducted, randomized and controlled clinical trial is still the gold standard in obtaining valid causal effects of medical products, such designs can be either impractical or too burdensome to conduct in pre-market and post-market studies. For example, very frequently a prospective, controlled cohort design is used for the Post-approval Study (PAS) of medical devices at the post-market phase. Average causal effect (ACE) estimation methods for these non-randomized studies have typically relied on standard propensity scoring techniques, which assumes all covariates are measured accurately, no measurement and no unobserved factors influence the treatment and outcome (Rosenbaum and Rubin, 1983,1984; Dehejia and Wahba, 1999). However, covariates are often measured with unobservable error. Ignoring measurement error in covariates (a fairly common issue in medical and public health research) may lead to misleading inferences on average treatment or exposure effects evaluation in non-randomized studies. So, this R package is developed resulting from our causal methodology research on extending the standard causal inference framework to allow covariate measurement error and developing EM algorithm for ACE estimation in a likelihood based approach, to enhance the evaluation of the safety and efficacy of medical products (including devices, drugs, and biologics) in both the pre-market and post-market phases. Notice that the current outcome setting for this package is continuous with Gaussian distribution. For non-Gaussian continuous outcomes, they can be converted to Gaussian before running it. This package will be extended to include categorical outcomes soon in 2013. More methodology details are discussed in the technical report listed below.

Acknowledgement: This research is sponsored by FDA Critical Path Grants.

Note: This R package (software) is only for internal review for noq, not for public circulation or use, since the manuscript supporting this newly developed method is still under revision which is to be submitted to *Journal of the Royal Statistical Society, series B (Statistical Methodology)* in spring 2013. Once this manuscript is accepted and published in a suitable journal after peer review, we will make this R-package available in public for circulation and use as FDA regulation allowed.

References

Yi Huang, Karen Bandeen-Roche, Xiaoyu Dong, Andrew Raim, Constantine Frangakis, Cunlin Wang. Average Causal Effect Estimation Allowing Covariate Measurement Error Using Latent Propensity Score. Technical Report, 2012.

Examples

```

data(breastpump)
y <- breastpump$tylall
z <- breastpump$bpregbin

w <- breastpump$pirlevel
sigma.e <- pi/sqrt(3)
ratio <- 1.5
sigma.w <- sigma.e * sqrt(1 + ratio^2) / ratio
sw.log <- (log(w) - mean(log(w))) / sigma.w

xa <- model.matrix(~ p9 + bwt + bfeddur + maritalc + white + parity + retwk + smokem3 + bmi + boy + gage +
as.factor(educr) + as.factor(typedeliv), data = breastpump)

X <- cbind(xa[,-1], sw.log)
p <- ncol(X)

# Try J = 5 classes
J <- 5

# EM initialization
# alpha0.0 should be ordered from smallest to largest. Length of alpha0.0 should be J-1
beta1.0 <- c(0.2,0.1,0.1,0.2,0.1)
beta2.0 <- c(0,0,0,0,0)
sigma.0 <- sqrt(var(y))
lambda.0 <- c(0.1,0.1,0.2,0.4,0.1)
alpha0.0 <- c(0,0.05,0.10,0.15)
alpha.0 <- c(0.01,0.05,0.05,-0.1,0.01,0.05,-0.15,0.1,0.02,0.01,-0.01,0.02,0.03,0.04,-0.03,0.02,0.01)

em.result <- EM(y = y, z = z, beta1 = beta1.0, beta2 = beta2.0, lambda = lambda.0,
alpha0 = alpha0.0, alpha = alpha.0, sigma = sigma.0, X = X, J = J,
debug = FALSE, maxiter = 3000, debug.nr = FALSE, maxiter.nr = 1,
tol = 1e-6)

V <- cov.louis(y, z, em.result$beta1, em.result$beta2, em.result$lambda,
em.result$alpha0, em.result$alpha, em.result$sigma, X, J)

ace.res <- new.ace(V, y, z, em.result$beta1, em.result$beta2, em.result$lambda,
em.result$alpha0, em.result$alpha, em.result$sigma, X, J)

print(ace.res)

```

breastpump

Study on Breast Pump usage

Description

A subsample of 1000 observations taken from a study on Breast Pump usage

Format

A data frame with 1000 rows and 17 variables

p9 Maternal age

bwt Birth weight

pirlevel Poverty level, measured with error

ty1all A comorbidity score, on the log scale

bpregbin Indicator for regular breast pump usage

bfeddur Breast feeding duration, in weeks

maritalc Marital status

educr Education

white Indicator for race

parity Indicator for having one or more children

retwk Indicator for returning to work within the first year

smokem3 Smoking status at third month postpartum

bmi BMI before pregnancy

bmiclass Categorical BMI before pregnancy

boy Gender of infant

gage Gestational age

typedeliv Type of delivery

Source

TBD

cov.louis

Estimate Covariance Matrix for EM estimate by Louis Method

Description

Estimate covariate for EM estimate by Louis (1982) method.

Usage

cov.louis(y, z, beta1, beta2, lambda, alpha0, alpha, sigma, X, J)

Arguments

| | |
|--------|--|
| y | Continuous outcome (n dimension vector) |
| z | Treatment vector (n dimension vector) |
| beta1 | Outcome effects for treated (J dimension vector) |
| beta2 | Outcome effects for untreated (J dimension vector) |
| lambda | Treatment effects (J dimension vector) |
| alpha0 | Ordered intercepts for latent regression model (J-1 dimension) |
| alpha | Slopes for latent regression model (p dimension) |
| sigma | Standard deviation for normal outcome |
| X | $n \times p$ covariate for latent regression model |
| J | Number of latent groups |

Value

Covariance matrix for EM estimate

References

Louis, T. A. Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society. Series B, (44)2:226-233, 1982.

d.group

Complete Data Density

Description

Complete Data Density for the ACE mixture model, when groups (η_1, \dots, η_n) are observed.

Usage

d.group(y, z, eta, beta1, beta2, lambda, sigma, log = FALSE)

Arguments

| | |
|--------|--|
| y | Continuous outcome (n dimension vector) |
| z | Treatment vector (n dimension vector) |
| eta | Vector of group labels (n dimension vector) |
| beta1 | Outcome effects for treated (J dimension vector) |
| beta2 | Outcome effects for untreated (J dimension vector) |
| lambda | Treatment effects (J dimension vector) |
| sigma | Standard deviation for normal outcome |
| log | TRUE if log-density is to be returned, FALSE otherwise |

Value

Density value

| | |
|-----------|--------------------------------------|
| d.mixture | <i>Density for ACE mixture model</i> |
|-----------|--------------------------------------|

Description

Density for the ACE mixture model where Pi matrix has already been computed.

Usage

```
d.mixture(y, z, Pi, beta1, beta2, lambda, sigma, log = FALSE)
```

Arguments

| | |
|--------|---|
| y | Continuous outcome (n dimension vector) |
| z | Treatment vector (n dimension vector) |
| Pi | $n \times J$ matrix giving prior probabilities that i th observation is in j th group |
| beta1 | Outcome effects for treated (J dimension vector) |
| beta2 | Outcome effects for untreated (J dimension vector) |
| lambda | Treatment effects (J dimension vector) |
| sigma | Standard deviation for normal outcome |
| log | TRUE if log-density is to be returned, FALSE otherwise |

Value

Density value

| | |
|-------------|--------------------------------------|
| d.mixture.x | <i>Density for ACE mixture model</i> |
|-------------|--------------------------------------|

Description

Density for the ACE mixture model

Usage

```
d.mixture.x(y, z, alpha0, alpha, X, beta1, beta2, lambda, sigma, log = FALSE)
```

Arguments

| | |
|--------|--|
| y | Continuous outcome (n dimension vector) |
| z | Treatment vector (n dimension vector) |
| alpha0 | Ordered intercepts for latent regression model (J-1 dimension) |
| alpha | Slopes for latent regression model (p dimension) |
| X | $n \times p$ covariate for latent regression model |
| beta1 | Outcome effects for treated (J dimension vector) |
| beta2 | Outcome effects for untreated (J dimension vector) |
| lambda | Treatment effects (J dimension vector) |
| sigma | Standard deviation for normal outcome |
| log | TRUE if log-density is to be returned, FALSE otherwise |

Value

Density value

EM

EM Algorithm for ACE finite mixture model

Description

This function carries out the EM algorithm to compute the MLE for the ACE finite mixture likelihood with continuous outcome and binary treatment indicator. This likelihood assumes a latent regression model which clusters observations into one of J groups.

Usage

```
EM(y, z, beta1, beta2, lambda, alpha0, alpha, sigma, X, J, debug = FALSE,
  maxiter = Inf, tol = 1e-08, method = "NR", debug.nr = FALSE,
  tol.nr = tol, maxiter.nr = 1)
```

Arguments

| | |
|--------|---|
| y | Continuous outcome (n dimension vector) |
| z | Treatment vector (n dimension vector) |
| beta1 | Initial guess of outcome effects for treated (J dimension vector) |
| beta2 | Initial guess of outcome effects for untreated (J dimension vector) |
| lambda | Initial guess of treatment effects (J dimension vector) |
| alpha0 | Initial guess of ordered intercepts for latent regression model (J-1 dimension) |
| alpha | Initial guess of slopes for latent regression model (p dimension) |
| sigma | Initial guess of standard deviation for normal outcome |
| X | $n \times p$ covariate for latent regression model |

| | |
|------------|---|
| J | Number of latent groups |
| debug | If TRUE, print extra debug information |
| maxiter | Maximum number of EM iterations |
| tol | Convergence tolerance for EM |
| method | Use "NR" for Newton-Raphson iterations to compute next (alpha0, alpha). Or use "optim" to compute them with R's optimizer directly. |
| debug.nr | If TRUE, print extra debug information within Newton-Raphson iterations |
| tol.nr | Convergence tolerance for internal Newton-Raphson iterations |
| maxiter.nr | Maximum iterations for internal Newton-Raphson |

Details

The EM iterations for beta1, beta2, lambda, and sigma all have a closed form, but the iterations for (alpha0, alpha) must be computed numerically. The default method is to use One-step Newton Raphson (method = "NR"), but the user may instead choose iterative Newton-Raphson by setting tol.nr and maxiter.nr, or use optim by setting method = "optim".

Convergence statuses are:

- 0 if Converged
- 1 if "negative" progress was made after 75 iterations
- 2 if log-likelihood value of NA or NaN was obtained
- 3 if failed to converge within maxiter iterations

Value

| | |
|-------------|--|
| beta1 | Estimate for β_1 |
| beta2 | Estimate for β_2 |
| lambda | Estimate for λ |
| alpha0 | Estimate for α_0 |
| alpha | Estimate for α |
| sigma | Estimate for σ |
| loglik | log-likelihood corresponding to MLE |
| status | An integer indicating the convergence status |
| message | A message with the convergence status |
| tol | Convergence tolerance reached (difference between current and last log-likelihood) |
| elapsed.sec | Elapsed time in seconds to run EM |
| method | The method used to fit (α_0, α) |
| itr | Number of EM iterations |

| | |
|----------------|-----------------------------------|
| FormHessian.Q3 | <i>Hessian of the Q3 function</i> |
|----------------|-----------------------------------|

Description

Form the Hessian of the Q3 function, which is the EM log-likelihood with respect to (alpha0, alpha)

Usage

```
FormHessian.Q3(alpha0, alpha, Post, X, J)
```

Arguments

| | |
|--------|---|
| alpha0 | Initial guess of ordered intercepts for latent regression model (J-1 dimension) |
| alpha | Initial guess of slopes for latent regression model (p dimension) |
| Post | $n \times J$ matrix of posterior probabilities that i th subject is in j th group |
| X | $n \times p$ covariate for latent regression model |
| J | Number of latent groups |

Value

| | |
|---|----------------|
| H | Hessian matrix |
|---|----------------|

| | |
|--------------|---|
| FormScore.Q3 | <i>Score Vector for (alpha0, alpha)</i> |
|--------------|---|

Description

Form the Score Vector for (alpha0, alpha)

Usage

```
FormScore.Q3(alpha0, alpha, Post, X, J)
```

Arguments

| | |
|--------|---|
| alpha0 | Initial guess of ordered intercepts for latent regression model (J-1 dimension) |
| alpha | Initial guess of slopes for latent regression model (p dimension) |
| Post | $n \times J$ matrix of posterior probabilities that i th subject is in j th group |
| X | $n \times p$ covariate for latent regression model |
| J | Number of latent groups |

Value

| | |
|---|--------------|
| S | Score vector |
|---|--------------|

naive.ace

*Naive ACE estimator***Description**

Compute Naive ACE estimator by propensity score matching

Usage

```
naive.ace(y, z, xa, w, J)
```

Arguments

| | |
|----|---|
| y | Continuous outcome (n dimension vector) |
| z | Treatment vector (n dimension vector) |
| xa | Covariates which have measured accurately |
| w | Covariates which have measured with error |
| J | Number of groups to use for propensity score matching |

Value

| | |
|---------|---------------------------------|
| ace.est | ACE estimate |
| ace.se | Standard error for ACE estimate |

new.ace

*New ACE estimator and its standard error***Description**

Compute New ACE estimator and its standard error

Usage

```
new.ace(cov.mle, y, z, beta1, beta2, lambda, alpha0, alpha, sigma, X, J)
```

Arguments

| | |
|---------|---|
| cov.mle | An estimate for the covariance matrix of the MLE of $(\beta_1, \beta_2, \lambda, \alpha_0, \alpha, \sigma)$ |
| y | Continuous outcome (n dimension vector) |
| z | Treatment vector (n dimension vector) |
| beta1 | Outcome effects for treated (J dimension vector) |
| beta2 | Outcome effects for untreated (J dimension vector) |
| lambda | Treatment effects (J dimension vector) |

| | |
|--------|--|
| alpha0 | Ordered intercepts for latent regression model (J-1 dimension) |
| alpha | Slopes for latent regression model (p dimension) |
| sigma | Standard deviation for normal outcome |
| X | $n \times p$ covariate for latent regression model |
| J | Number of latent groups |

Value

| | |
|---------|---------------------------------|
| ace.est | ACE estimate |
| ace.se | Standard error for ACE estimate |

NewtonRaphson.Q3 *Iterative Newton-Raphson procedure for (alpha0, alpha)*

Description

Carries out iterative Newton-Raphson procedure for (alpha0, alpha),

Usage

```
NewtonRaphson.Q3(alpha0, alpha, Post, X, J, tol = 1e-05, debug = FALSE,
  maxiter = Inf)
```

Arguments

| | |
|---------|---|
| alpha0 | Initial guess of ordered intercepts for latent regression model (J-1 dimension) |
| alpha | Initial guess of slopes for latent regression model (p dimension) |
| Post | $n \times J$ matrix of posterior probabilities that i th subject is in j th group |
| X | $n \times p$ covariate for latent regression model |
| J | Number of latent groups |
| tol | Convergence tolerance for EM |
| debug | If TRUE, print extra debug information |
| maxiter | Maximum number of EM iterations |

Value

| | |
|--------|--|
| alpha0 | Estimate for α_0 |
| alpha | Estimate for α |
| tol | Convergence tolerance reached (difference between current and last log-likelihood) |
| iter | Number of EM iterations |
| H | Hessian matrix from the last iteration |

NewtonRaphson.Q3.OneStep

One Step Newton-Raphson procedure for (alpha0, alpha)

Description

Carries out one step Newton-Raphson procedure for (alpha0, alpha). No need to check any convergence criteria here. This is a special case of NewtonRaphson.Q3 below

Usage

```
NewtonRaphson.Q3.OneStep(alpha0, alpha, Post, X, J, debug = FALSE)
```

Arguments

| | |
|--------|---|
| alpha0 | Initial guess of ordered intercepts for latent regression model (J-1 dimension) |
| alpha | Initial guess of slopes for latent regression model (p dimension) |
| Post | $n \times J$ matrix of posterior probabilities that i th subject is in j th group |
| X | $n \times p$ covariate for latent regression model |
| J | Number of latent groups |
| debug | If TRUE, print extra debug information |

Value

| | |
|--------|--|
| alpha0 | Estimate for α_0 |
| alpha | Estimate for α |
| H | Hessian matrix from the last iteration |

Pi.matrix

Density for ACE mixture model

Description

Density for the ACE mixture model

Usage

```
Pi.matrix(alpha0, alpha, X)
```

Arguments

| | |
|--------|--|
| alpha0 | Ordered intercepts for latent regression model (J-1 dimension) |
| alpha | Slopes for latent regression model (p dimension) |
| X | $n \times p$ covariate for latent regression model |

Value

$n \times J$ matrix giving probabilities that i th observation is in j th group

| | |
|----|--|
| Q3 | <i>EM log-likelihood with respect to (alpha0, alpha)</i> |
|----|--|

Description

The EM log-likelihood with respect to (alpha0, alpha)

Usage

Q3(theta, Post, X)

Arguments

| | |
|-------|---|
| theta | Vector of (alpha0, alpha) |
| Post | $n \times J$ matrix of posterior probabilities that i th subject is in j th group |
| X | $n \times p$ covariate for latent regression model |

Value

Q3 function value evaluated at theta

| | |
|-----------|---|
| r.mixture | <i>Generate observations from the ACE mixture model</i> |
|-----------|---|

Description

Generate observations directly from the ACE mixture model.

Usage

r.mixture(n, Pi, beta1, beta2, lambda, sigma)

Arguments

| | |
|--------|---|
| n | Number of observations to generate |
| Pi | $n \times J$ matrix giving probabilities that i th observation is in j th group |
| beta1 | Outcome effects for treated (J dimension vector) |
| beta2 | Outcome effects for untreated (J dimension vector) |
| lambda | Treatment effects (J dimension vector) |
| sigma | Standard deviation for normal outcome |

| Value | |
|-------|----------------------|
| y | Vector of outcomes |
| z | Vector of treatments |

Index

*Topic **ACE**

- d.group, [5](#)
- d.mixture, [6](#)
- d.mixture.x, [6](#)
- naive.ace, [10](#)
- new.ace, [10](#)
- r.mixture, [13](#)

*Topic **EM**

- cov.louis, [4](#)
- EM, [7](#)
- Q3, [13](#)

*Topic **Expectation-Maximization**

- EM, [7](#)

*Topic **Hessian**

- FormHessian.Q3, [9](#)

*Topic **Louis**

- cov.louis, [4](#)

*Topic **Newton-Raphson**

- FormHessian.Q3, [9](#)
- NewtonRaphson.Q3, [11](#)
- NewtonRaphson.Q3.OneStep, [12](#)

*Topic **Q3**

- FormScore.Q3, [9](#)

*Topic **Score**

- FormScore.Q3, [9](#)

*Topic **covariance**

- cov.louis, [4](#)

*Topic **datasets**

- breastpump, [3](#)

*Topic **iterative**

- NewtonRaphson.Q3, [11](#)

*Topic **mixture**

- d.group, [5](#)
- d.mixture, [6](#)
- d.mixture.x, [6](#)
- r.mixture, [13](#)

*Topic **one-step**

- NewtonRaphson.Q3.OneStep, [12](#)

*Topic **probabilities**

- Pi.matrix, [12](#)

- breastpump, [3](#)

- cov.louis, [4](#)

- d.group, [5](#)
- d.mixture, [6](#)
- d.mixture.x, [6](#)

- EM, [7](#)

- EstACEwErrorinX-package, [2](#)

- FormHessian.Q3, [9](#)

- FormScore.Q3, [9](#)

- naive.ace, [10](#)

- new.ace, [10](#)

- NewtonRaphson.Q3, [11](#)

- NewtonRaphson.Q3.OneStep, [12](#)

- Pi.matrix, [12](#)

- Q3, [13](#)

- r.mixture, [13](#)