Package: ATE.ERROR (via r-universe)

Title Estimating ATE with Misclassified Outcomes and Mismeasured Covariates

Version 1.0.0

Description Addressing measurement error in covariates and misclassification in binary outcome variables within causal inference, the 'ATE.ERROR' package implements inverse probability weighted estimation methods proposed by Shu and Yi (2017, <doi:10.1177/0962280217743777>; 2019, <doi:10.1002/sim.8073>). These methods correct errors to accurately estimate average treatment effects (ATE). The package includes two main functions: ATE.ERROR.Y() for handling misclassification in the outcome variable and ATE.ERROR.XY() for correcting both outcome misclassification and covariate measurement error. It employs logistic regression for treatment assignment and uses bootstrap sampling to calculate standard errors and confidence intervals, with simulated datasets provided for practical demonstration.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.2.3

Depends R (>= 2.10)

LazyData true

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Imports ggplot2, MASS, mvtnorm, rlang, stats

Author Aryan Rezanezhad [aut, cre], Grace Y. Yi [aut]

Maintainer Aryan Rezanezhad <Aryan.rzn@gmail.com>

Date/Publication 2024-09-10 09:10:10 UTC

Repository https://aryanrzn.r-universe.dev

RemoteUrl https://github.com/cran/ATE.ERROR

RemoteRef HEAD

RemoteSha 463ca12476a7d95a45a7b249618f9fdb43bcc7dc

Contents

ATE.ERROR.XY	*ATE.ERROR.XY Function for Estimating Average Treatment Effect (ATE) with Measurement Error in X and Misclassification in Y*

Description

The `ATE.ERROR.XY` function implements a method for estimating the Average Treatment Effect (ATE) that accounts for both measurement error in covariates and misclassification in the binary outcome variable Y.

Usage

```
ATE.ERROR.XY(
  Y_star,
  A,
  Z,
  X_star,
  p11,
  p10,
  sigma_epsilon,
  B = 100,
  Lambda = seq(0, 2, by = 0.5),
  extrapolation = "linear",
  bootstrap_number = 250
)
```

Arguments

Y_star	Numeric vector. The observed binary outcome variable, possibly misclassified.
A	Numeric vector. The treatment indicator (1 if treated, 0 if control).
Z	Numeric vector. A precisely measured covariate vector.
X_star	Numeric vector. A covariate vector subject to measurement error.
p11	Numeric. The probability of correctly classified Y given Y = 1.

p10	Numeric. The probability of misclassified Y given Y = 0.
sigma_epsilon	Numeric. The covariance matrix Sigma_epsilon for the measurement error model.
B	Integer. The number of simulated datasets.
Lambda	Numeric vector. A sequence of lambda values for simulated datasets.
extrapolation	Character. A regression model used for extrapolation ("linear", "quadratic", "nonlinear").
bootstrap_number	
	Numeric. The number of bootstrap samples (default is 250).

Details

The `ATE.ERROR.XY` function is designed to handle measurement error in covariates and misclassification in outcomes by using the augmented simulation-extrapolation approach.

Value

A list containing:

summary A data frame with the following columns:

- `Naive_ATE`: Naive estimate of the ATE.
- `Sigma_epsilon`: The covariance matrix Sigma_epsilon for the measurement error model.
- `p10`: The probability of misclassified Y given Y = 0.
- `p11`: The probability of correctly classified Y given Y = 1.
- `Extrapolation`: A regression model used for extrapolation ("linear", "quadratic", "nonlinear").
- `ATE`: Mean ATE estimate from the bootstrap samples.
- `SE`: Standard error of the ATE estimate.
- `CI`: 95% confidence interval for the ATE estimate.

boxplot A ggplot object representing the boxplot of the ATE estimates.

Examples

```
library(ATE.ERROR)
data(Simulated_data)
Y_star <- Simulated_data$Y_star
A <- Simulated_data$T
Z <- Simulated_data$Z
X_star <- Simulated_data$X_star
p11 <- 0.8
p10 <- 0.2
sigma_epsilon <- 0.1
B <- 100
Lambda <- seq(0, 2, by = 0.5)
bootstrap_number <- 10
result <- ATE.ERROR.XY(Y_star, A, Z, X_star, p11, p10, sigma_epsilon, B, Lambda,
                       "linear", bootstrap_number)
print(result$summary)
print(result$boxplot)
```

ATE.ERROR.Y   *ATE.ERROR.Y Function for Estimating Average Treatment Effect (ATE) with Misclassification in Y*

Description

This function performs estimation of the Average Treatment Effect (ATE) using the ATE.ERROR.Y method, which accounts for misclassification in the binary outcome variable Y. The method calculates consistent estimates of the ATE in the presence of misclassified outcomes by leveraging logistic regression and bootstrap sampling.

Usage

```
ATE.ERROR.Y(Y_star, A, Z, X, p11, p10, bootstrap_number = 250)
```

Arguments

Y_star	Numeric vector. The observed binary outcome variable, which may be subject to misclassification.
A	Numeric vector. The binary treatment indicator (1 if treated, 0 if control).
Z	Numeric vector. A precisely measured covariate vector.
X	Numeric vector. A precisely measured covariate vector.
p11	Numeric. The probability of correctly classified Y given Y = 1.
p10	Numeric. The probability of misclassified Y given Y = 0.
bootstrap_number	
	Integer. The number of bootstrap samples (default is 250) used to obtain the associated variance estimate.

Details

The function first calculates consistent estimates of the ATE, correcting for misclassification in the outcome variable Y. The logistic model is used to estimate the propensity scores for the treatment assignment, which are then adjusted using the provided misclassification probabilities p11 and p10. Bootstrap sampling is performed to estimate the variance and construct confidence intervals for the ATE estimates.

Value

A list containing:

summary A data frame with the following columns:

- **Naive_ATE**: Naive estimate of the ATE, ignoring misclassification.
- **ATE**: Mean ATE estimate from the bootstrap samples, accounting for misclassification.

- **SE**: Standard error of the ATE estimate.
- **CI**: 95% confidence interval for the ATE estimate.

boxplot A ggplot object representing the boxplot of the ATE estimates.

Examples

```
library(ATE.ERROR)
data(Simulated_data)
Y_star <- Simulated_data$Y_star
A <- Simulated_data$T
Z <- Simulated_data$Z
X <- Simulated_data$X
p11 <- 0.8
p10 <- 0.2
bootstrap_number <- 250
result <- ATE.ERROR.Y(Y_star, A, Z, X, p11, p10, bootstrap_number)
print(result$summary)
print(result$boxplot)
```

Naive_Estimation *Naive Estimation of Average Treatment Effect*

Description

This function performs a naive estimation of the ATE. This approach gives us the so-called "naive estimate" by ignoring the difference between (X*, Y*) and (X, Y).

Usage

```
Naive_Estimation(Y_star, A, Z, X_star)
```

Arguments

Y_star	A numeric vector of outcomes with potential misclassification.
A	A numeric vector of treatment assignments.
Z	A numeric vector of covariate Z.
X_star	A numeric vector of covariate X with measurement error.

Value

A numeric value representing the estimated treatment effect.

Examples

```
library(ATE.ERROR)
data(Simulated_data)
Y_star <- Simulated_data$Y_star
A <- Simulated_data$T
Z <- Simulated_data$Z
X_star <- Simulated_data$X_star
Naive_ATE_XY <- Naive_Estimation(Y_star, A, Z, X_star)
print(Naive_ATE_XY)
```

Simulated_data   *Simulated Data*

Description

A dataset containing simulated data generated by the generate_data function. This data includes misclassified outcome Y*, treatment assignment T, and covariates X and Z.

Usage

```
Simulated_data
```

Format

A data frame with 5000 rows and 6 variables:

X a numeric vector generated from a standard normal distribution (mean = 0, standard deviation = 1)

X_star a numeric vector where X_star is equal to X plus a random error. The random error is generated from a normal distribution with mean 0 and standard deviation 0.1

Y a numeric vector generated from a Bernoulli distribution with a probability depending on T, Z, and X

Y_star a numeric vector where Y_star is generated from a binomial distribution depending on Y with probabilities 0.8 if Y equals 1 and 0.2 if Y equals 0

T a numeric vector generated from a binomial distribution with probability calculated using the logistic function of the sum of 0.2, Z, and X

Z a numeric vector generated from a standard normal distribution (mean = 0, standard deviation = 1)

Source

Shu D, Yi GY (2019). Weighted causal inference methods with mismeasured covariates and misclassified outcomes. Statistics in Medicine. 38:1835-1854. doi:10.1002/sim.8073

True_Estimation   *True Estimation of Average Treatment Effect*

Description

This function performs a true estimation of the Average Treatment Effect (ATE) using the generated values for X and Y. The consistent estimator is calculated as the difference between the expected value of the outcome for the treated group and the expected value of the outcome for the control group.

Usage

```
True_Estimation(Y, A, Z, X)
```

Arguments

Y	A numeric vector of outcomes.
A	A numeric vector of treatment assignments.
Z	A numeric vector of covariate Z.
X	A numeric vector of covariate X.

Details

The expected value for the treated group, $E(Y\_1)$, is calculated as the mean of the product of the treatment assignment and the outcome divided by the estimated propensity score.

The expected value for the control group, $E(Y\_0)$, is calculated as the mean of the product of the control assignment and the outcome divided by one minus the estimated propensity score.

The propensity score is estimated by applying a logistic regression model to the true values of the covariates and treatment assignments.

Value

A numeric value representing the estimated treatment effect.

Examples

```
library(ATE.ERROR)
data(Simulated_data)
Y <- Simulated_data$Y
A <- Simulated_data$T
Z <- Simulated_data$Z
X <- Simulated_data$X
True_ATE <- True_Estimation(Y, A, Z, X)
print(True_ATE)
```

Index