Package 'ui'

June 19, 2019

Title Uncertainty Intervals and Sensitivity Analysis for Missing Data

Version 0.1.0

Author Minna Genbaeck [aut, cre],

Maintainer Minna Genbaeck <minna.genback@umu.se>

Description Implements functions to derive uncertainty intervals for (i) regression (linear and probit) parameters when outcome is missing not at random (non-ignorable missingness) introduced in Genbaeck, M., Stanghellini, E., de Luna, X. (2015) <doi:10.1007/s00362-014-0610-x> and Genbaeck, M., Ng, N., Stanghellini, E., de Luna, X. (2018) <doi:10.1007/s10433-017-0448-x>; and (ii) double robust and outcome regression estimators of average causal effects (on the treated) with possibly unobserved confounding introduced in Genbaeck, M., de Luna, X. (2018) <doi:10.1111/biom.13001>.

Depends R (>= 3.5)

Imports Matrix, maxLik, mvtnorm, numDeriv, graphics, stats

Suggests MASS

Encoding UTF-8

LazyData true

License GPL-2

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-06-19 12:10:03 UTC

R topics documented:

`gridrho.f` *Support function for* `ui.causal`

Description

Divides the `rho` interval into a grid

Usage

```
gridrho.f(rho, gridn, rho.plotrange, plot)
```

Arguments

`rho`	interval that should be divided
`gridn`	number of gridpoints
`rho.plotrange`	a larger interval of grids to be used in a plot
`plot`	whether or not the larger interval of grids should be created

grr *Gradient for the loglikelihood used by ui.probit*

Description

This function derives the gradient in order for `ui.probit` to run faster.

Usage

```
grr(par, rho, X.z = X.z, X.y = X.y, y = y, z = z)
```

Arguments

`par`	Coefficients.
`rho`	Rho.
`X.z`	Covariate matrix for missingness.
`X.y`	Covariate matrix for outcome.
`y`	Outcome.
`z`	Missing or not.

hess *Hessian for the loglikelihood used by ui.probit*

Description

This function derives the hessian in order for `ui.probit` to run faster.

Usage

```
hess(par, rho, X.z = X.z, X.y = X.y, y = y, z = z)
```

Arguments

`par`	Coefficients.
`rho`	Rho.
`X.z`	Covariate matrix for missingness.
`X.y`	Covariate matrix for outcome.
`y`	Outcome.
`z`	Missing or not.

interv.p *Print interval in parantesis*

Description

This function allows you to print an interval (vector of two elements) in a parantesis single element.

Usage

```
interv.p(v, digits = 3)
```

Arguments

v	Lower and upper bounds.
digits	Number of decimals.

lambda0 *Inverse Mills rato*

Description

This function allows you to calculate the inverse Mills ratio.

Usage

```
lambda0(x)
```

Arguments

x	Vector

lambda1 *Inverse Mills rato*

Description

This function allows you to calculate the inverse Mills ratio.

Usage

```
lambda1(x)
```

Arguments

x	Vector

LogL.probit *Loglikelihood used by ui.probit*

Description

This function derives the Loglikelihood for `ui.probit`.

Usage

```
LogL.probit(par, rho, X.z = X.z, X.y = X.y, y = y, z = z)
```

Arguments

par	Coeficient values the logliklihood should be drived at.
rho	The value of the sensitivity parameter.
X.z	covariate matrix for missingness mechanism
X.y	covariate matrix for the outcome regression
y	outcome vector
z	indicator of wether y is missing or not

Logl.sandACT *Loglikelohood used in sandwich estimator of average causal effect on the treated for DR*

Description

Loglikelohood used in sandwich estimator of average causal effect on the treated for DR, support function for `ui.causal`

Usage

```
Logl.sandACT(x, X, z)
```

Arguments

x	coefficents.
X	Covariate matrix.
z	Missing or not.

ML.probit *Fit maximum likelihood for fixed values of rho*

Description

This is a support function for ui.probit

Usage

```
ML.probit(out.formula, mis.formula = NULL, data, rho = c(-0.5, 0.5),
  progress = TRUE, method = "NR")
```

Arguments

out.formula	Formula for outcome regression.
mis.formula	Formula for regression model for the missingness mechanism.
data	Data frame containing the variables in the formulas
rho	Vector containing the values of rho for which we want to fit the likelihood.
progress	If TRUE prints out process time for each maximazation of the likelihood.
method	Maximazation method to be passed through maxLik

plot.uicausal *Plot of UI and CI*

Description

Plot function for objects returned from ui.causal. Plots confidence intervals for different values of rho and the uncertainty interval.

Usage

```
## S3 method for class 'uicausal'
plot(x, DR = TRUE, main = "", xlab = NULL,
  ylab = "", ...)
```

Arguments

x	An object of class uicausal
DR	If TRUE the doubly robust estimator is plotted, otherwise the outcome regression estimator is plotted.
main	Main title, default is no title.
xlab	Title for xaxis, default is expression(rho).
ylab	Title for y axis, default is no title.
...	Additional arguments, use is discouraged.

plot.uiols *Plot of UI and CI*

Description

Plot function for objects returned from `ui.ols`. Plots confidence intervals, coefficients and significans assuming ignorability and the uncertainty interval under non-ignorability.

Usage

```
## S3 method for class 'uiols'
plot(x, plot.all = TRUE, which = NA,
  intercept = FALSE, ylab = NULL, col = c("black", "red"), ...)
```

Arguments

`x`	An object of class uiols
`plot.all`	If TRUE, plots all covariates.
`which`	Specify which variables should be plotted by either sending in their names in a vector or a vector with their numbers (1 intercept, 2 for the first covariate etc.).
`intercept`	If TRUE, also plots the intercept.
`ylab`	Vector of names for the y-axis, default is the variable names.
`col`	Vector containing the color of confidence intervals (default black) and uncertainty intervals (default red).
`...`	Additional arguments, use is discouraged.

plot.uiprobit *Plot of UI and CI*

Description

Plot function for objects returned from `ui.probit`. Plots confidence intervals, coefficients and significans assuming ignorability and the uncertainty interval under non-ignorability.

Usage

```
## S3 method for class 'uiprobit'
plot(x, plot.all = TRUE, which = NA,
  intercept = FALSE, ylab = NULL, col = c("black", "red"), ...)
```

Arguments

x	An object of class uiprobit
plot.all	If TRUE, plots all covariates.
which	Specify which variables should be plotted by either sending in their names in a vector or a vector with their numbers (1 for the first covariate, 2 for the second etc.).To plot the intercept, set intercept as TRUE.
intercept	If TRUE, also plots the intercept.
ylab	Vector of names for the y-axis, default is the variable names.
col	Vector containing the color of confidence intervals (default black) and uncertainty intervals (default red).
...	Additional arguments, use is discouraged.

---

print.uicausal *Print function for object of class uicausal*

Description

Print function for object of class uicausal

Usage

```
## S3 method for class 'uicausal'
print(x, digits = 3, digitsci = digits,
  digitsui = digits, ...)
```

Arguments

x	An object of returned from ui.causal
digits	number of digits to be printed.
digitsci	number of digits to be printed in the confidence interval.
digitsui	number of digits to be printed in the uncertainty interval.
...	Additional arguments, use is discouraged.

print.uiols *Prints objects of class uiols*

Description

Prints objects of class uiols

Usage

```
## S3 method for class 'uiols'
print(x, digits = 3, digitsci = digits,
  digitsui = digits, ...)
```

Arguments

x	an objects returned from ui.ols
digits	number of digits to be printed.
digitsci	number of digits to be printed in the confidence interval.
digitsui	number of digits to be printed in the uncertainty interval.
...	Additional arguments, use is discouraged.

print.uiprobit *Prints objects of class uiprobit*

Description

Prints objects of class uiprobit

Usage

```
## S3 method for class 'uiprobit'
print(x, digits = 3, digitsci = digits,
  digitsui = digits, ...)
```

Arguments

x	an objects returned from ui.probit
digits	number of digits to be printed.
digitsci	number of digits to be printed in the confidence interval.
digitsui	number of digits to be printed in the uncertainty interval.
...	Additional arguments, use is discouraged.

profile.uicausal *Plot of UI and CI*

Description

Plot function for objects returned from `ui.causal`. Plots confidence intervals for different values of rho0=rho1=rho.

Usage

```
## S3 method for class 'uicausal'
profile(fitted, DR = TRUE, main = "", xlab = NULL,
  ylab = "", ...)
```

Arguments

`fitted`	An object of class uicausal
`DR`	If TRUE, plots both DR if FALSE OR.
`main`	Main title, default is no title.
`xlab`	Title for x-axis, default is `expression(rho)`.
`ylab`	Title for y-axis, default is the variable names.
`...`	Additional arguments, use is discouraged.

profile.uiols *Plot of UI and CI*

Description

Plot function for objects returned from `ui.ols`. Plots confidence intervals for different values of rho and the uncertainty interval.

Usage

```
## S3 method for class 'uiols'
profile(fitted, plot.all = TRUE, which = NA,
  intercept = FALSE, xlab = NULL, ylab = NULL, ...)
```

Arguments

fitted	An object of class uiols
plot.all	If TRUE, plots all covariates.
which	Specify which variables should be plotted by either sending in their names in a vector or a vector with their numbers (1 intercept, 2 for the first covariate etc.).
intercept	If TRUE, also plots the intercept.
xlab	Title for x-axis, default is expression(rho).
ylab	Title for y-axis, default is the variable names.
...	Additional arguments, for instance margins.

---

profile.uiprobit *Plot of UI and CI*

Description

Plot function for objects returned from ui.probit. Plots confidence intervals for different values of rho and the uncertainty interval.

Usage

```
## S3 method for class 'uiprobit'
profile(fitted, plot.all = TRUE, which = NA,
  intercept = FALSE, xlab = NULL, ylab = NULL, cex.lab = 2,
  mar = c(6, 6, 2, 2), ...)
```

Arguments

fitted	An object of class uiprobit
plot.all	If TRUE, plots all covariates.
which	Specify which variables should be plotted by either sending in their names in a vector or a vector with their numbers (1 intercept, 2 for the first covariate etc.).
intercept	If TRUE, also plots the intercept.
xlab	Title for x-axis, default is expression(rho).
ylab	Title for y-axis, default is the variable names.
cex.lab	Size of lables.
mar	Margin around panels in plot.
...	Additional arguments, use is discouraged.

sandACT — *Calculates standard error of Average causal effect on the treated*

Description

This is a support function for `ui.causal` and calculates standard error of Average causal effect on the treated for the doubly robust estimator.

Usage

```
sandACT(deltasigma1, X, Xz, y, z, u, BetaOLSy0, phat, NaivEst, n1, n0, N,
  p, pz)
```

Arguments

`deltasigma1`	Coefficients.
`X`	Covariate matrix outcome.
`Xz`	Covariate matrix treatment.
`y`	Outcome vector.
`z`	Missingness indicator.
`u`	Fitted values from propensity score regression.
`BetaOLSy0`	Coefficients from non-treated regression
`phat`	Fitted propensity scores.
`NaivEst`	Naiv estimates.
`n1`	Number of treated.
`n0`	Number of non-treated.
`N`	Total number.
`p`	Number of covariates outcome regression.
`pz`	Number of covariates treatment regression.

sandImpACE — *Calculates standard error of Average causal effect*

Description

This is a support function for `ui.causal` and calculates standard error of Average causal effect for the regression imputation estimator.

Usage

```
sandImpACE(X, y, z, BetaOLSy0, BetaOLSy1, NaivEst, N, p)
```

Arguments

X	Covariate matrix.
y	Outcome vector.
z	missingness indicator.
BetaOLSy0	Coefficients from non-treated regression.
BetaOLSy1	Coefficients from treated regression.
NaivEst	Naiv estimates.
N	Total number.
p	Number of covariates outcome regression.

---

sandImpACT *Calculates standard error of Average causal effect on the treated*

Description

This is a support function for ui.causal and calculates standard error of Average causal effect on the treated for the regression imputation estimator.

Usage

```
sandImpACT(X, y, z, BetaOLSy0, NaivEst, n1, N, p)
```

Arguments

X	Covariate matrix.
y	Outcome vector.
z	missingness indicator
BetaOLSy0	Coefficients from non-treated regression
NaivEst	Naiv estimates.
n1	Number of treated.
N	Total number.
p	Number of covariates outcome regression.

se.ols *Calculation of se for OLS*

Description

This function calculates the se for UI based on OLS when we have MNAR data, for `ui.ols`.

Usage

```
se.ols(X, sigmaOLScor, u, gridrho)
```

Arguments

`X`	Covariate matrix.
`sigmaOLScor`	Output from sigmaOLScor1
`u`	Fitted values from mis.model.
`gridrho`	Values of rho.

sigmaOLScor0 *Correction of OLS sigma for causal effects*

Description

This function is a bias correction of the residual standard deviation under MNAR, for `ui.causal`.

Usage

```
sigmaOLScor0(X, sigmaOLS, n, p, u, gridrho)
```

Arguments

`X`	Covariate matrix outcome.
`sigmaOLS`	Residual sd from outcome regression.
`n`	Number of complete cases.
`p`	Number of covariates outcome regression.
`u`	Fitted values from propensity score regression.
`gridrho`	Values of rho.

sigmaOLScor1 *Correction of OLS sigma*

Description

This function is a bias correction of the residual standard deviation under MNAR, used by `ui.causal` and `ui.ols`.

Usage

```
sigmaOLScor1(X, sigmaOLS, n, p, u, gridrho)
```

Arguments

`X`	Covariate matrix outcome.
`sigmaOLS`	Residual sd from outcome regression.
`n`	Number of complete cases.
`p`	Number of covariates outcome regression.
`u`	Fitted values from propensity score regression.
`gridrho`	Values of rho.

ui.causal *Uncertainty intervals for Average Causal Effects*

Description

This function allows you to derive uncertainty intervals for the average causal effect (ACE) or the average causal effect on the treated (ACT). The function uses a regression imputation estimator and a doubly robust estimator. The uncertainty intervals can be used as a sensitivity analysis to unconfoundedness. Note that `rho=0` render the same results as assuming no unobserved confounding.

Usage

```
ui.causal(out.formula, treat.formula, data, rho = c(-0.3, 0.3),
  rho0 = NULL, rho1 = NULL, ACT = FALSE, sand = TRUE, gridn = 21,
  plot = TRUE, rho.plotrange = c(-0.5, 0.5), alpha = 0.05)
```

Arguments

out.formula	Formula for the outcome regression models
treat.formula	Formula for the propensity score model (regression model for treatment assignment).
data	data.frame containing the variables in the formula.
rho	Pre-specified interval for rho0 and rho1.
rho0	Pre-specified value of rho0, if an interval it has to be the same as rho1.
rho1	Pre-specified value of rho1, if an interval it has to be the same as rho0.
ACT	If TRUE Average Causal effect of the Treated is calculated, if FALSE Average Causal effect is calculated. Default is FALSE.
sand	Specifies which estimator of the standard errors should be used for OR, see details.
gridn	Number of fixed points within the rho interval for which sigma0 and sigma1 should be estimated.
plot	If TRUE the function runs slightly slower but you will be able to plot your results using plot.uicausal.
rho.plotrange	an interval larger than rho for the plot using plot.uicausal.
alpha	Default 0.05 corresponding to a confidence level of 95 for CI and UI.

Details

In order to visualize the results, you can use plot.uicausal. Details about estimators can be found in Genbäck and de Luna (2018)

The standard errors are calculated with the following estimators:

DR ACE - simplified sandwich estimator

DR ACT - sandwich estimator

OR ACE - if sand=TRUE sandwich estimator (default and recommended), if sand=FALSE large sample variance

OR ACT - if sand=TRUE sandwich estimator (default and recommended), if sand=FALSE large sample variance

Value

A list containing:

call	The matched call
rho0	The rage of rho0 from which the ui is calculated
rho1	If ACT==FALSE,range of rho1 from which the ui is calculated
out.model0	Outcome regression model for non-treated.
out.model1	Outcome regression model for treated.
treat.model	Regression model for treatment mechanism (propensity score).
sigma0	Consistent estimate of sigma0 for different values of rho0

sigma1	Consistent estimate of sigma1 for different values of rho1
DR	DR inference, confidence intervals for different pre-specified values of rho for the OR estimator, uncertainty interval, coefficient estimates, confounding bias, indentification interval, standard error etc.
OR	OR inference, confidence intervals for different pre-specified values of rho for the OR estimator, uncertainty interval, coefficient estimates, confounding bias, indentification interval, standard error etc.

Author(s)

Minna Genbäck

References

Genbäck, M., de Luna, X. (2018). Causal Inference Accounting for Unobserved Confounding after Outcome Regression and Doubly Robust Estimation. *Biometrics*. DOI: 10.1111/biom.13001

Examples

```
library(MASS)
n<-500
delta<-c(-0.3,0.65)
rho<-0.3
X<-cbind(rep(1,n),rnorm(n))
x<-X[,-1]
s0<-2
s1<-3
error<-mvrnorm(n, c(0,0,0), matrix(c(1,0.6,0.9,0.6,4,0.54,0.9,0.54,9), ncol=3))
zstar<-X%*%delta+error[,1]
z<-as.numeric(zstar>0)
y1<-ifelse(x< (-1),0.2*x-0.1*x^2, ifelse(x< 1,0.3*x, ifelse(x<3,0.4-0.1*x^2,-0.2-0.1*x)))+error[,3]
y0<-ifelse(x<1.5, x-0.4*x^2, ifelse(x<2, -0.15-0.25*x+0.5*x^2, 1.85-0.25*x))+error[,2]
y<-y0
y[z==1]<-y1[z==1]
data<-data.frame(y,z,x)


ui<-ui.causal(y~x, z~x, data=data, rho=c(0,0.3), ACT=FALSE)
ui
plot(ui)
profile(ui)
mean(y1-y0)

ui<-ui.causal(y~x, z~x, data=data, rho=c(0,0.3), ACT=TRUE)
ui
plot(ui)
mean(y1[z==1]-y0[z==1])
```

ui.ols *Uncertainty intervals for OLS regression*

Description

This function allows you to derive uncertainty intervals for OLS regression when there is missing data in the continuous outcome. The uncertainty intervals can be used as a sensitivity analysis to ignorability (missing at random). Note that rho=0 render the same results as a complete case analysis.

Usage

```
ui.ols(out.formula, mis.formula = NULL, data, rho = c(-0.3, 0.3),
  alpha = 0.05, gridn = 101)
```

Arguments

out.formula	Formula for outcome regression.
mis.formula	Formula for missingness mechanism. If NULL the same covariates as in the outcome regression will be used.
data	data.frame containing the variables in the formula.
rho	The limits of rho for which the uncertainty interval should be constructed.
alpha	Default 0.05 corresponding to a confidence level of 95 for CI and UI.
gridn	The number of distinct points within the interval rho at which confidence intervals should be constructed. Default is 101.

Details

In order to visualize the results, you can use plot.uiols, or profile.uiols.

Value

A list containing:

call	The matched call
ci	Confidence intervals for different values of rho
ui	Uncertainty intervals
coef	Estimated coefficients (outcome regression) for different values of rho
out.model	Outcome regression model when rho=0.
mis.model	Regression model for missingness mechanism (selection).
rho	The range of rho for which we want to construct an uncertainty interval
gridrho	The values of rho for which bias and standard errors are derived
sigma	Consistant estimate of sigma
se	Standard error for different values of rho
ciols	Confidence intervals from a complete case analysis
ident.bound	Bounds for the coefficient estimates.

Author(s)

Minna Genbäck

References

Genbäck, M., Stanghellini, E., de Luna, X. (2015). Uncertainty Intervals for Regression Parameters with Non-ignorable Missingness in the Outcome. *Statistical Papers*, 56(3), 829-847.

Examples

```
library(MASS)
n<-500
delta<-c(0.5,0.3,0.1)
beta<-c(0.8,-0.2,0.3)
X<-cbind(rep(1,n),rnorm(n),rbinom(n,1,0.5))
x<-X[,-1]
rho=0.4
error<-mvrnorm(n,c(0,0),matrix(c(1,rho*2,rho*2,4),2))
zstar<-X%*%delta+error[,1]
z<-as.numeric(zstar>0)
y<-X%*%beta+error[,2]
y[z==0]<-NA
data<-data.frame(y,x,z)
ui<-ui.ols(y~X1+X2,data=data,rho=c(-0.5,0.5))
ui
plot(ui)
```

---

ui.probit *Uncertainty intervals for probit regression*

Description

This function allows you to derive uncertainty intervals for probit regression when there is missing data in the binary outcome. The uncertainty intervals can be used as a sensitivity analysis to ignorability (missing at random), and are derived by maximum likelihood. Note that `rho=0` render the same results as a complete case analysis.

Usage

```
ui.probit(out.formula, mis.formula = NULL, data, rho = c(-0.3, 0.3),
  progress = TRUE, max.grid = 0.1, alpha = 0.05, method = "NR")
```

Arguments

`out.formula`	Formula for outcome regression.
`mis.formula`	Formula for missingness mechanism. If NULL the same covariates as in the outcome regression will be used.

data	data.frame containing the variables in the formula.
rho	Vector containing the values of rho for which we want to fit the likelihood.
progress	If TRUE prints out process time for each maximization of the likelihood.
max.grid	Maximum distance between two elements in rho, if two wide there can difficulties with convergence of the maximum likelihood.
alpha	Default 0.05 corresponding to a confidence level of 95 for CI and UI.
method	Maximization method to be passed through maxLik

Details

In order to visualize the results, you can use plot.uiprobit or profile.uiprobit.

Value

A list containing:

coef	Estimated coefficients (outcome regression) for different values of rho.
rho	The values of rho for which the likelihood is maximized.
vcov	Covariance matrix.
ci	Confidence intervals for different values of rho.
ui	Uncertainty intervals.
out.model	Outcome regression model when rho=0.
mis.model	Regression model for missingness mechanism (selection).
se	Standard errors from outcome regression.
value	Value of maximum likelihood for different values of rho.
y	Outcome vector.
z	Indicator variable of observed outcome.
X.y	Covariate matrix for outcome regression.
X.z	Covariate matrix for missingness mechanism (selection regression model).
max.info	Information about the maximization procedure. Includes whether it converged, message, method and number of iterations.

Author(s)

Minna Genbäck

References

Genbäck, M., Ng, N., Stanghellini, E., de Luna, X. (2018). Predictors of Decline in Self-reported Health: Addressing Non-ignorable Dropout in Longitudinal Studies of Aging. *European journal of ageing*, 15(2), 211-220.

Examples

```
library(MASS)
n<-500

delta<-c(0.5,0.6,0.1,-1,1)
beta<-c(-0.3,-0.5,0,-0.4,-0.3)

X<-cbind(rep(1,n),rnorm(n),runif(n),rbinom(n,2,0.5),rbinom(n,1,0.5))
x<-X[,-1]
rho=0.4
error<-mvrnorm(n,c(0,0),matrix(c(1,rho,rho,1),2))

zstar<-X%*%delta+error[,1]
z<-as.numeric(zstar>0)

ystar<-X%*%beta+error[,2]
y<-as.integer(ystar>0)
y[z==0]<-NA
data=data.frame(y=y,x1=x[,1],x2=x[,2],x3=x[,3],x4=x[,4])

m<-ui.probit(y~x1+x2+x3+x4,data=data,rho=c(0,0.5))
m
plot(m)
profile(m)
```

Index