

Package ‘tmle’

March 25, 2015

Type Package

Title Targeted Maximum Likelihood Estimation

Version 1.2.0-4

Date 2014-3-4

Author Susan Gruber, in collaboration with Mark van der Laan

Maintainer Susan Gruber <sgruber@cal.berkeley.edu>

Copyright Copyright 2012. The Regents of the University of California (Regents). All Rights Reserved.

Depends SuperLearner

Description tmle implements targeted maximum likelihood estimation, first described in van der Laan and Rubin, 2006 (Targeted Maximum Likelihood Learning, The International Journal of biostatistics, 2(1), 2006. This version adds the tmleMSM function to the package, for estimating the parameters of a marginal structural model (MSM) for a binary point treatment effect. The tmle function calculates the adjusted marginal difference in mean outcome associated with a binary point treatment, for continuous or binary outcomes. Relative risk and odds ratio estimates are also reported for binary outcomes. Missingness in the outcome is allowed, but not in treatment assignment or baseline covariate values. Effect estimation stratified by a binary mediating variable is also available. The population mean is calculated when there is missingness, and no variation in the treatment assignment. An ID argument can be used to identify repeated measures. Default settings call SuperLearner to estimate the Q and g portions of the likelihood, unless values or a user-supplied regression function are passed in as arguments.

License BSD_3_clause + file LICENSE | GPL-2

URL <http://cran.r-project.org/package=tmle>

NeedsCompilation no

Repository CRAN

Date/Publication 2014-03-09 08:00:09

R topics documented:

tmle-package 2

calcParameters	3
calcSigma	4
estimateG	5
estimateQ	6
fev	8
SL.glm.interaction	8
summary.tmle	9
summary.tmleMSM	11
tmle	12
tmleMSM	16
tmleNews	20

Index	21
--------------	-----------

tmle-package	<i>Targeted Maximum Likelihood Estimation with Super Learning</i>
--------------	---

Description

Targeted maximum likelihood estimation of marginal treatment effect of a binary point treatment on a continuous or binary outcome, adjusting for baseline covariates. Missingness in the outcome is accounted for in the estimation procedure. The population mean outcome is calculated when there is missingness and no treatment. Controlled direct effect estimation is available, and MSM parameter estimation for binary point treatment effects. Optional data-adaptive estimation of Q and g portions of the likelihood using the SuperLearner package is strongly encouraged.

Details

Package: tmle
 Type: Package
 Version: 1.2.0-3
 Date: 2012-11-15
 License: BSD | GPL-2

Author(s)

Susan Gruber, in collaboration with Mark van der Laan.
 Maintainer: Susan Gruber, <sgruber@cal.berkeley.edu>

References

1. Gruber, S. and van der Laan, M.J. (2012), tmle: An R Package for Targeted Maximum Likelihood Estimation. *Journal of Statistical Software*, 51(13), 1-35. <http://www.jstatsoft.org/v51/i13/>

2. Gruber, S. and van der Laan, M.J. (2009), Targeted Maximum Likelihood Estimation: A Gentle Introduction. *U.C. Berkeley Division of Biostatistics Working Paper Series*. Working Paper 252. <http://www.bepress.com/ucbbiostat/paper252>
3. Gruber, S. and van der Laan, M.J. (2010), A Targeted Maximum Likelihood Estimator of a Causal Effect on a Bounded Continuous Outcome. *The International Journal of Biostatistics*, 6(1), 2010.
4. Rosenblum, M. and van der Laan, M.J. (2010). Targeted Maximum Likelihood Estimation of the Parameter of a Marginal Structural Model. *The International Journal of Biostatistics*, 6(2), 2010.
5. van der Laan, M.J. and Rubin, D. (2006), Targeted Maximum Likelihood Learning. *The International Journal of Biostatistics*, 2(1). <http://www.bepress.com/ijb/vol2/iss1/11/>
6. van der Laan, M.J., Rose, S., and Gruber, S., editors, (2009) Readings in Targeted Maximum Likelihood Estimation . *U.C. Berkeley Division of Biostatistics Working Paper Series*. Working Paper 254. <http://www.bepress.com/ucbbiostat/paper254>

See Also

[tmle](#), [tmleMSM](#)

calcParameters

Calculate Parameter Estimates (calcParameters)

Description

An internal function called by the `tmle` function to calculate the population mean effect when there is missingness in the data, but no treatment assignment. When observations are in treatment and control groups, estimates the additive treatment effect, and if the outcome is binary, also the relative risk and odds ratio parameters. P-values and 95% confidence intervals are also calculated (on the log scale for RR and OR).

Usage

```
calcParameters(Y, A, I.Z, Delta, g1W, g0W, Q, mu1, mu0, id, family)
```

Arguments

Y	continuous or binary outcome variable
A	binary treatment indicator, 1 - treatment, 0 - control
I.Z	Indicator $Z=z$, needed for CDE estimation
Delta	indicator of missing outcome. 1 - observed, 0 - missing
g1W	censoring mechanism estimates, $P(A = 1 W) * P(Delta = 1 A, W)$
g0W	censoring mechanism estimates, $P(A = 0 W) * P(Delta = 1 A, W)$
Q	a 3-column matrix ($Q(A, W)$, $Q(1, W)$, $Q(0, W)$)
mu1	targeted estimate of $E(Y A = 1, W)$
mu0	targeted estimate of $E(Y A = 0, W)$

id subject identifier
 family family specification for regressions, generally ‘gaussian’ for continuous outcomes,
 ‘binomial’ for binary outcomes

Value

EY1 Population mean outcome estimate, variance, p-value, 95% confidence interval
 (missingness only, no treatment assignment), or NULL
 ATE additive treatment effect estimate, variance, p-value, 95% confidence interval,
 or NULL
 RR relative risk estimate, p-value, 95% confidence interval, log(RR), variance(log(RR)),
 or NULL
 OR odds ratio estimate, p-value, 95% confidence interval, log(OR), variance(log(OR)),
 or NULL

Author(s)

Susan Gruber

See Also

[tmle](#), [estimateQ](#), [estimateG](#), [tmleMSM](#), [calcSigma](#)

calcSigma	<i>Calculate Variance-Covariance Matrix for MSM Parameters (calc-Sigma)</i>
-----------	---

Description

An internal function called by the `tmleMSM` function to calculate the variance-covariance matrix of the parameter estimates based on the influence curve of the specified MSM.

Usage

```
calcSigma(hAV, gAVW, Y, Q, mAV, covar.MSM, covar.MSMA0, covar.MSMA1, I.V,  
Delta, ub, id, family)
```

Arguments

hAV values used in numerator of weights applied to the estimation procedure
 gAVW $p(A = a|V, W, T) * p(Delta = 1|A, V, W, T)$
 Y continuous or binary outcome variable
 Q estimated $P(Y|A, V, W, T, Delta = 1)$, typically targeted values Q^* are passed
 in
 mAV predicted values for *EY1* from the MSM using the targeted estimates for *psi*

covar.MSM	covariate values used as predictors for the MSM when A=a
covar.MSMA0	covariate values used as predictors for the MSM when A=0
covar.MSMA1	covariate values used as predictors for the MSM when A=1
I.V	indicator that observation is in stratum of interest
Delta	indicator of missing outcome. 1 - observed, 0 - missing
ub	upper bound on weights
id	subject identifier
family	'gaussian' for continuous outcomes, 'binomial' for binary outcomes

Value

sigma	influence-curve based variance-covariance matrix. See Rosenblum&vanderLaan2010 for details.
-------	---

Author(s)

Susan Gruber

See Also

[tmle](#), [estimateQ](#), [estimateG](#), [tmleMSM](#)

estimateG

Estimate Treatment or Missingness Mechanism

Description

An internal function called by the `tmle` function to obtain an estimate of conditional treatment assignment probabilities $P(A = 1|W)$, and conditional probabilities for missingness, $P(Delta = 1|A, W)$. The estimate can be based on user-supplied values, a user-supplied regression formula, or a data-adaptive super learner fit. If the SuperLearner package is not available, and there are no user-specifications, estimation is carried out using main terms regression with `glm`. These main terms-based estimates may yield poor results.

Usage

```
estimateG(d, g1W, gform, SL.library, id, verbose, message, outcome, newdata=d)
```

Arguments

d	dataframe with binary dependent variable in the first column, predictors in remaining columns
g1W	vector of values for $P(A = 1 W)$, $P(Z = 1 A, W)$, or $P(Delta = 1 Z, A, W)$
gform	regression formula of the form $A \sim W1$, (dependent variable is one of A, Z, D) if specified this overrides the call to SuperLearner

SL.library	vector of prediction algorithms used by SuperLearner, default value is ('SL.glm', 'SL.step', 'SL.glm.interaction')
id	subject identifier
verbose	status messages printed if set to TRUE
message	text specifies whether treatment or missingness mechanism is being estimated
outcome	A, D, Z to indicate which quantity is being estimated.
newdata	optional dataset to be used for prediction after fitting on d.

Value

g1W	a vector containing values for $P(A = 1 W)$, matrix for $P(Z = 1 A, W)$, evaluated at $A=0$, $A=1$, or matrix $P(Delta = 1 Z, A, W)$ evaluated at $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$
coef	coefficients for each term in the working model used for estimation if glm was used
type	estimation procedure

Author(s)

Susan Gruber

See Also

[tmle](#), [estimateQ](#), [calcParameters](#), [tmleMSM](#), [calcSigma](#)

estimateQ

Initial Estimation of Q portion of the Likelihood

Description

An internal function called by the tmle function to obtain an initial estimate of the Q portion of the likelihood based on user-supplied matrix values for predicted values of (counterfactual outcomes) $Q(0, W)$, $Q(1, W)$, or a user-supplied regression formula, or based on a data-adaptively selected SuperLearner fit. In the absence of user-supplied values, a user-supplied regression formula takes precedence over data-adaptive super-learning.

Usage

```
estimateQ(Y, Z, A, W, Delta, Q, Qbounds, Qform, maptoYstar, SL.library, cvQinit,
family, id, verbose)
```

Arguments

Y	continuous or binary outcome variable
Z	optional binary indicator for intermediate covariate for controlled direct effect estimation
A	binary treatment indicator, 1 - treatment, 0 - control
W	vector, matrix, or dataframe containing baseline covariates
Delta	indicator of missing outcome. 1 - observed, 0 - missing
Q	3-column matrix ($Q(A, W)$, $Q(0, W)$, $Q(1, W)$)
Qbounds	Bounds on predicted values for Q, set to alpha for logistic fluctuation, or range(Y) if not user-supplied
Qform	regression formula of the form $Y \sim A + W$
maptoYstar	if TRUE indicates continuous Y values should be shifted and scaled to fall between (0,1)
SL.library	specification of prediction algorithms, default is ('SL.glm', 'SL.step', 'SL.glm.interaction'). In practice, including more prediction algorithms in the library improves results.
cvQinit	logical, whether or not to estimate cross-validated values for initial Q, default=FALSE
family	family specification for regressions, generally 'gaussian' for continuous outcomes, 'binomial' for binary outcomes
id	subject identifier
verbose	status message printed if set to TRUE

Value

Q	$n \times 3$ matrix, columns contain the initial estimate of $[Q(A, W) = E(Y A = a, W), Q(0, W) = E(Y A = 0, W), Q(1, W) = E(Y A = 1, W)]$. For controlled direct estimation, $n \times 5$ matrix, $E(Y Z, A, W)$, evaluated at $(z, a), (0, 0), (0, 1), (1, 0), (1, 1)$ on scale of linear predictors
Qfamily	'binomial' for targeting with logistic fluctuation, 'gaussian' for linear fluctuation
coef	coefficients for each term in working model used for initial estimation of Q if glm used.
type	type of estimation procedure

Author(s)

Susan Gruber

See Also[tmle](#), [estimateG](#), [calcParameters](#), [tmleMSM](#), [calcSigma](#)

fev	<i>Forced Expiratory Volume (FEV) Data (fev)</i>
-----	--

Description

Sample of 654 youths, aged 3 to 19, in the area of East Boston during middle to late 1970's. Interest concerns the relationship between smoking and FEV. Since the study is necessarily observational, statistical adjustment via regression models clarifies the relationship.

Usage

```
data(fev)
```

Format

A data frame with 654 observations on the following 5 variables.

age a numeric vector
 fev a numeric vector
 ht a numeric vector
 sex a numeric vector
 smoke a numeric vector

Source

Kahn M (2005). An Exhalent Problem for Teaching Statistics. The Journal of Statistical Education, 13(2).

Rosner, B. (1999), Fundamentals of Biostatistics, 5th Ed., Pacific Grove, CA: Duxbury.

SL.glm.interaction	<i>Wrapper function for SuperLearner prediction algorithm</i>
--------------------	---

Description

A function in the default SuperLearner library used by tmle. This wrapper is built into SuperLearner 2.0-6, and is defined here for backwards-compatibility with older versions of SuperLearner. This function may be incorporated into any user-supplied SuperLearner library, but will not ordinarily be called directly by the user.

Usage

```
SL.glm.interaction(Y.temp, X.temp, newX.temp, family, obsWeights, ...)
```


Arguments

Y.temp	The outcome
X.temp	The training data set
newX.temp	The validation data set (the observations to make predictions from)
family	outcome distribution, either gaussian or binomial
obsWeights	observation weights, currently unused
...	additional arguments

Value

out	The predicted outcome on the scale of the outcome
fit	A list of elements needed for the predict method on the specific algorithm

Author(s)

Susan Gruber

summary.tmlc

Summarization of the results of a call to the tmlc routine

Description

These functions are all [methods](#) for class tmlc, tmlc.list, summary.tmlc, summary.tmlc.list objects

Usage

```
## S3 method for class 'tmlc'
summary(object, ...)
## S3 method for class 'tmlc.list'
summary(object, ...)
## S3 method for class 'tmlc'
print(x, ...)
## S3 method for class 'tmlc.list'
print(x, ...)
## S3 method for class 'summary.tmlc'
print(x, ...)
## S3 method for class 'summary.tmlc.list'
print(x, ...)
```

Arguments

object	an object of class tmlc or tmlc.list.
x	an object of class tmlc or tmlc.list for summary functions, class summary.tmlc or summary.tmlc.list for print functions.
...	currently ignored.

Details

`print.tmle` prints the estimate, variance, p-value, and 95% confidence interval only. `print.summary.tmle`, called indirectly by entering the command `summary(result)` (where `result` has class `tmle`), outputs additional information. Controlled direct effect estimates have class `tmle.list`, a list of two objects of class `tmle`. The first item corresponds to $Z = 0$, the second to $Z = 1$

Value

<code>estimates</code>	list of parameter estimates, pvalues, and 95% confidence intervals
<code>Qmodel</code>	working model used to obtain initial estimate of Q portion of the likelihood, if <code>glm</code> used
<code>Qterms</code>	terms in the model for Q
<code>Qcoef</code>	coefficient of each term in model for Q
<code>gmodel</code>	model used to estimate treatment mechanism g
<code>gterms</code>	terms in the treatment mechanism model
<code>gcoef</code>	coefficient of each term in model for treatment mechanism
<code>gtype</code>	description of estimation procedure for treatment mechanism, e.g. "SuperLearner"
<code>g.Zmodel</code>	model used to estimate intermediate variable assignment mechanism g.Z
<code>g.Zterms</code>	terms in the intermediate mechanism model
<code>g.Zcoef</code>	coefficient of each term in model for intermediate mechanism
<code>g.Ztype</code>	description of estimation procedure for intermediate variable
<code>g.Deltamodel</code>	model used to estimate missingness mechanism g.Delta
<code>g.Deltaterms</code>	terms in the missingness mechanism model
<code>g.Deltacoef</code>	coefficient of each term in model for missingness mechanism
<code>g.Deltatype</code>	description of estimation procedure for missingness

Author(s)

Susan Gruber

See Also

[tmle](#)

Examples

```
# generate data
set.seed(10)
n <- 500
W <- matrix(rnorm(n*3), ncol=3)
A <- rbinom(n,1, 1/(1+exp(-(.1*W[,1] - .1*W[,2] + .5*W[,3])))
Y <- A + 2*W[,1] + W[,3] + W[,2]^2 + rnorm(n)
colnames(W) <- paste("W",1:3, sep="")

result <- tmle(Y,A,W, Qform="Y~A+W1", g1W=rep(.5, n))
summary(result)
```

summary.tmlMSM *Summarization of the results of a call to the tmlMSM function*

Description

These functions are all [methods](#) for class tmlMSM, summary.tmlMSM objects

Usage

```
## S3 method for class 'tmlMSM'
summary(object, ...)
## S3 method for class 'tmlMSM'
print(x, ...)
## S3 method for class 'summary.tmlMSM'
print(x, ...)
```

Arguments

object	an object of class tmlMSM.
x	an object of class tmlMSM for summary functions, class summary.tmlMSM for print functions.
...	currently ignored.

Details

print.tmlMSM prints the estimate, standard error, p-value, and 95% confidence interval only. print.summary.tmlMSM, called indirectly by entering the command summary(result) (where result has class tmlMSM), outputs additional information.

Value

estimates	matrix of MSM parameter estimates, standard errors, pvalues, upper and lower bounds on 95% confidence intervals
sigma	variance-covariance matrix
Qmodel	working model used to obtain initial estimate of Q portion of the likelihood, if glm used
Qterms	terms in the model for Q
Qcoef	coefficient of each term in model for Q
gmodel	model used to estimate treatment mechanism g
gterms	terms in the treatment mechanism model
gcoef	coefficient of each term in model for treatment mechanism
gtype	description of estimation procedure for treatment mechanism, e.g. "SuperLearner"
g.AVmodel	model used to estimate $h(A,V)$ (or $h(A,T)$)
g.AVterms	terms in the model for $h(A,V)$

g.AVcoef	coefficient of each term in model for $h(A,V)$
g.AVtype	description of estimation procedure for $h(A,V)$
g.Deltamodel	model used to estimate missingness mechanism $g.Delta$
g.Deltaterms	terms in the missingness mechanism model
g.Deltacoef	coefficient of each term in model for missingness mechanism
g.Deltatype	description of estimation procedure for missingness
psi.Qinit	MSM parameter estimates based on initial (untargeted) estimated Q

Author(s)

Susan Gruber

See Also[tmleMSM](#)

tmle

*Targeted Maximum Likelihood Estimation***Description**

Targeted maximum likelihood estimation of parameters of a marginal structural model, and of marginal treatment effects of a binary point treatment on an outcome. In addition to the additive treatment effect, risk ratio and odds ratio estimates are reported for binary outcomes. The `tmle` function is generally called with arguments (Y, A, W) , where Y is a continuous or binary outcome variable, A is a binary treatment variable, ($A=1$ for treatment, $A=0$ for control), and W is a matrix or dataframe of baseline covariates. The population mean outcome is calculated when there is no variation in A . If values of binary mediating variable Z are supplied, estimates are returned at each level of Z . Missingness in the outcome is accounted for in the estimation procedure if missingness indicator Δ is 0 for some observations. Repeated measures can be identified using the `id` argument.

Usage

```
tmle(Y, A, W, Z=NULL, Delta = rep(1,length(Y)), Q = NULL, Q.Z1 = NULL, Qform = NULL,
     Qbounds = NULL, Q.SL.library = c("SL.glm", "SL.step", "SL.glm.interaction"),
     cvQinit = FALSE, g1W = NULL, gform = NULL, gbound = 0.025, pZ1=NULL,
     g.Zform = NULL, pDelta1 = NULL, g.Deltaform = NULL,
     g.SL.library = c("SL.glm", "SL.step", "SL.glm.interaction"),
     family = "gaussian", fluctuation = "logistic", alpha = 0.995, id=1:length(Y),
     verbose = FALSE)
```

Arguments

Y	continuous or binary outcome variable
A	binary treatment indicator, 1 - treatment, 0 - control
W	vector, matrix, or dataframe containing baseline covariates
Z	optional binary indicator for intermediate covariate for controlled direct effect estimation
Delta	indicator of missing outcome or treatment assignment. 1 - observed, 0 - missing
Q	optional $n \times 2$ matrix of initial values for Q portion of the likelihood, $(E(Y A = 0, W), E(Y A = 1, W))$
Q.Z1	optional $n \times 2$ matrix of initial values for Q portion of the likelihood, $(E(Y Z = 1, A = 0, W), E(Y Z = 1, A = 1, W))$. (When specified, values for $E(Y Z = 0, A = 0, W), E(Y Z = 0, A = 1, W)$ are passed in using the Q argument
Qform	optional regression formula for estimation of $E(Y A, W)$, suitable for call to <code>glm</code>
Qbounds	vector of upper and lower bounds on Y and predicted values for initial Q. Defaults to the range of Y, widened by 10% of the min and max values.
Q.SL.library	optional vector of prediction algorithms to use for SuperLearner estimation of initial Q
cvQinit	logical, if TRUE, estimates cross-validated predicted values using discrete super learning, default=FALSE
g1W	optional vector of conditional treatment assignment probabilities, $P(A = 1 W)$
gform	optional regression formula of the form $A \sim W$, if specified this overrides the call to SuperLearner
gbound	value between (0,1) for truncation of predicted probabilities. See Details section for more information
pZ1	optional $n \times 2$ matrix of conditional probabilities $P(Z = 1 A = 0, W), P(Z = 1 A = 1, W)$
g.Zform	optional regression formula of the form $Z \sim A + W$, if specified this overrides the call to SuperLearner
pDelta1	optional matrix of conditional probabilities for missingness mechanism, $n \times 2$ when Z is NULL $P(Delta = 1 A = 0, W), P(Delta = 1 A = 1, W)$. $n \times 4$ otherwise, $P(Delta = 1 Z = 0, A = 0, W), P(Delta = 1 Z = 0, A = 1, W), P(Delta = 1 Z = 1, A = 0, W), P(Delta = 1 Z = 1, A = 1, W)$
g.Deltaform	optional regression formula of the form $Delta \sim A + W$, if specified this overrides the call to SuperLearner
g.SL.library	optional vector of prediction algorithms to use for SuperLearner estimation of <code>g1W</code> or <code>pDelta1</code>
family	family specification for working regression models, generally 'gaussian' for continuous outcomes (default), 'binomial' for binary outcomes
fluctuation	'logistic' (default), or 'linear'
alpha	used to keep predicted initial values bounded away from (0,1) for logistic fluctuation
id	optional subject identifier
verbose	status messages printed if set to TRUE (default=FALSE)

Details

g.bounds defaults to (0.025, 0.975) for treatment effect estimates. If only one value is provided, symmetric truncation levels are assumed. Bounds default to (0.025, 1) for estimating the population mean outcome.

W should only contain covariates that are factors when Super Learner is not used to estimate Q or g.

Controlled direct effects are estimated when binary covariate Z is non-null. The tmle function returns an object of class tmle.list, a list of two items of class tmle. The first corresponds to estimates obtained when Z is fixed at 0, the second corresponds to estimates obtained when Z is fixed at 1.

Q.SL.library defaults to ('SL.glm', 'SL.step', 'SL.glm.interaction')

g.SL.library Defaults to ('SL.glm', 'SL.step', 'SL.glm.interaction')

This choice is simply because these algorithms are included in the base R installation. See SuperLearner help files for further information.

Value

estimates	list with elements EY1 (population mean), ATE (additive treatment effect), RR (relative risk), OR (odds ratio). Each element in the estimates of these is itself a list containing <ul style="list-style-type: none"> • psi - parameter estimate • pvalue - two-sided p-value • CI - 95% confidence interval • var.psi - Influence-curve based variance of estimate (ATE parameter only) • log.psi - Parameter estimate on log scale (RR and OR parameters) • var.log.psi - Influence-curve based variance of estimate on log scale (RR and OR parameters)
Qinit	initial estimate of Q. Qinit\$coef are the coefficients for a glm model for Q, if applicable. Qinit\$Q is an $n \times 2$ matrix, where n is the number of observations. Columns contain predicted values for $Q(0, W)$, $Q(1, W)$ using the initial fit. Qinit\$type is method for estimating Q
Qstar	targeted estimate of Q, an $n \times 2$ matrix with predicted values for $Q(0, W)$, $Q(1, W)$ using the updated fit
g	treatment mechanism estimate. A list with three items: g\$g1W contains estimates of $P(A = 1 W)$ for each observation, g\$coef the coefficients for the model for g when glm used, g\$type estimation procedure
g.Z	intermediate covariate assignment estimate (when applicable). A list with three items: g.Z\$g1W an $n \times 2$ matrix containing values of $P(Z = 1 A = 1, W)$, $P(Z = 1 A = 0, W)$ for each observation, g.Z\$coef the coefficients for the model for g when glm used, g.Z\$type estimation procedure
g.Delta	missingness mechanism estimate. A list with three items: g.Delta\$g1W an $n \times 4$ matrix containing values of $P(Delta = 1 Z, A, W)$ for each observation, with $(Z=0, A=0)$, $(Z=0, A=1)$, $(Z=1, A=0)$, $(Z=1, A=1)$. (When Z is NULL, columns 3 and 4 are duplicates of 1 and 2.) g.Delta\$coef the coefficients for the model for g when glm used, g.Delta\$type estimation procedure

Author(s)

Susan Gruber <sgruber@cal.berkeley.edu>, in collaboration with Mark van der Laan.

References

1. Gruber, S. and van der Laan, M.J. (2012), tmle: An R Package for Targeted Maximum Likelihood Estimation. *Journal of Statistical Software*, 51(13), 1-35. <http://www.jstatsoft.org/v51/i13/>
2. Gruber, S. and van der Laan, M.J. (2009), Targeted Maximum Likelihood Estimation: A Gentle Introduction. *U.C. Berkeley Division of Biostatistics Working Paper Series*. Working Paper 252. <http://www.bepress.com/ucbbiostat/paper252>
3. Gruber, S. and van der Laan, M.J. (2010), A Targeted Maximum Likelihood Estimator of a Causal Effect on a Bounded Continuous Outcome. *The International Journal of Biostatistics*, 6(1), 2010.
4. Rosenblum, M. and van der Laan, M.J. (2010). Targeted Maximum Likelihood Estimation of the Parameter of a Marginal Structural Model. *The International Journal of Biostatistics*, 6(2), 2010.
5. van der Laan, M.J. and Rubin, D. (2006), Targeted Maximum Likelihood Learning. *The International Journal of Biostatistics*, 2(1). <http://www.bepress.com/ijb/vol2/iss1/11/>
6. van der Laan, M.J., Rose, S., and Gruber, S., editors, (2009) Readings in Targeted Maximum Likelihood Estimation. *U.C. Berkeley Division of Biostatistics Working Paper Series*. Working Paper 254. <http://www.bepress.com/ucbbiostat/paper254>

See Also

[summary.tmle](#), [estimateQ](#), [estimateG](#), [calcParameters](#), [tmleMSM](#), [calcSigma](#)

Examples

```
library(tmle)
set.seed(1)
n <- 250
W <- matrix(rnorm(n*3), ncol=3)
A <- rbinom(n,1, 1/(1+exp(-(.2*W[,1] - .1*W[,2] + .4*W[,3])))
Y <- A + 2*W[,1] + W[,3] + W[,2]^2 + rnorm(n)

# Example 1. Simplest function invocation
# SuperLearner called to estimate Q, g
# Delta defaults to 1 for all observations
result1 <- tmle(Y,A,W)
summary(result1)

# Example 2:
# User-supplied regression formulas to estimate Q and g
# binary outcome
# n <- 250
W <- matrix(rnorm(n*3), ncol=3)
colnames(W) <- paste("W",1:3, sep="")
A <- rbinom(n,1, plogis(0.6*W[,1] + 0.4*W[,2] + 0.5*W[,3]))
Y <- rbinom(n,1, plogis(A + 0.2*W[,1] + 0.1*W[,2] + 0.2*W[,3]^2 ))
result2 <- tmle(Y,A,W, family="binomial", Qform=Y~A+W1+W2+W3, gform=A~W1+W2+W3)
```

```

summary(result2)

# Example 3: Population mean outcome
# User-supplied (misspecified) model for Q,
# Super learner called to estimate g, g.Delta
# approx. 20% missing at random
Y <- W[,1] + W[,2]^2 + rnorm(n)
Delta <- rbinom(n, 1, 1/(1+exp(-(1.7-1*W[,1]))))
result3 <- tmle(Y,A=NULL,W, Delta=Delta, Qform="Y~A+W1+W2+W3")
print(result3)

# Example 4: Controlled direct effect
# User-supplied models for g, g.Z
A <- rbinom(n,1,.5)
Z <- rbinom(n, 1, plogis(.5*A + .1*W[,1]))
Y <- 1 + A + 10*Z + W[,1]+ rnorm(n)

CDE <- tmle(Y,A,W, Z, gform="A~1", g.Zform = "Z ~ A + W1")
total.effect <- tmle(Y,A, W, gform="A~1")
print(CDE)
print(total.effect)

```

tmleMSM

Targeted Maximum Likelihood Estimation of Parameter of MSM

Description

Targeted maximum likelihood estimation of the parameter of a marginal structural model (MSM) for binary point treatment effects. The `tmleMSM` function is minimally called with arguments (Y, A, W, MSM) , where Y is a continuous or binary outcome variable, A is a binary treatment variable, ($A=1$ for treatment, $A=0$ for control), and W is a matrix or dataframe of baseline covariates. MSM is a valid regression formula for regressing Y on any combination of A , V , W , T , where V defines strata and T represents the time at which repeated measures on subjects are made. Missingness in the outcome is accounted for in the estimation procedure if missingness indicator Δ is 0 for some observations. Repeated measures can be identified using the `id` argument.

Usage

```

tmleMSM(Y, A, W, V, T = rep(1,length(Y)), Delta = rep(1, length(Y)), MSM,
v = NULL, Q = NULL, Qform = NULL, Qbounds = c(-Inf, Inf),
Q.SL.library = c("SL.glm", "SL.step", "SL.glm.interaction"),
cvQinit = FALSE, hAV = NULL, hAVform = NULL, g1W = NULL,
gform = NULL, pDelta1 = NULL, g.Deltaform = NULL,
g.SL.library = c("SL.glm", "SL.step", "SL.glm.interaction"),
ub = 1/0.025, family = "gaussian", fluctuation = "logistic",
alpha = 0.995, id = 1:length(Y), inference = TRUE, verbose = FALSE)

```


Arguments

Y	continuous or binary outcome variable
A	binary treatment indicator, 1 - treatment, 0 - control
W	vector, matrix, or dataframe containing baseline covariates. Factors are not currently allowed.
V	vector, matrix, or dataframe of covariates used to define strata
T	optional time for repeated measures data
Delta	indicator of missing outcome or treatment assignment. 1 - observed, 0 - missing
MSM	MSM of interest, specified as valid right hand side of a regression formula (see examples)
v	optional value defining the strata of interest ($V = v$) for stratified estimation of MSM parameter
Q	optional $n \times 2$ matrix of initial values for Q portion of the likelihood, $(E(Y A = 0, W), E(Y A = 1, W))$
Qform	optional regression formula for estimation of $E(Y A, W)$, suitable for call to <code>glm</code>
Qbounds	vector of upper and lower bounds on Y and predicted values for initial Q
Q.SL.library	optional vector of prediction algorithms to use for SuperLearner estimation of initial Q
cvQinit	logical, if TRUE, estimates cross-validated predicted values using discrete super learning, default=FALSE
hAV	optional $n \times 2$ matrix used in numerator of weights for updating covariate and the influence curve. If unspecified, defaults to conditional probabilities $P(A = 1 V)$ or $P(A = 1 T)$, for repeated measures data. For unstabilized weights, pass in an $n \times 2$ matrix of all 1s
hAVform	optional regression formula of the form $A \sim V + T$, if specified this overrides the call to SuperLearner
g1W	optional vector of conditional treatment assignment probabilities, $P(A = 1 W)$
gform	optional regression formula of the form $A \sim W$, if specified this overrides the call to SuperLearner
pDelta1	optional $n \times 2$ matrix of conditional probabilities for missingness mechanism, $P(Delta = 1 A = 0, V, W, T), P(Delta = 1 A = 1, V, W, T)$.
g.Deltaform	optional regression formula of the form $Delta \sim A + W$, if specified this overrides the call to SuperLearner
g.SL.library	optional vector of prediction algorithms to use for SuperLearner estimation of g1W or pDelta1
ub	upper bound on observation weights. See Details section for more information
family	family specification for working regression models, generally 'gaussian' for continuous outcomes (default), 'binomial' for binary outcomes
fluctuation	'logistic' (default), or 'linear'

alpha	used to keep predicted initial values bounded away from (0,1) for logistic fluctuation
id	optional subject identifier
inference	if TRUE, variance-covariance matrix, standard errors, pvalues, and 95% confidence intervals are calculated. Setting to FALSE saves a little time when bootstrapping.
verbose	status messages printed if set to TRUE (default=FALSE)

Details

ub bounds the IC by bounding the factor $h(A, V)/[g(A, V, W)P(\Delta = 1|A, V, W)]$ between 0 and ub, default value = 1/0.025.

Q.SL.library Defaults to ('SL.glm', 'SL.step', 'SL.glm.interaction')

g.SL.library Defaults to ('SL.glm', 'SL.step', 'SL.glm.interaction')

This choice is simply because these algorithms are included in the base R installation. See SuperLearner help files for further information.

Value

psi	MSM parameter estimate
sigma	variance covariance matrix
se	standard errors extracted from sigma
pvalue	two-sided p-value
lb	lower bound on 95% confidence interval
ub	upper bound on 95% confidence interval
epsilon	fitted value of epsilon used to target initial Q
psi.Qinit	MSM parameter estimate based on untargeted initial Q
Qstar	targeted estimate of Q, an $n \times 2$ matrix with predicted values for $Q(0, W)$, $Q(1, W)$ using the updated fit
Qinit	initial estimate of Q. Qinit\$coef are the coefficients for a glm model for Q, if applicable. Qinit\$Q is an $n \times 2$ matrix, where n is the number of observations. Columns contain predicted values for $Q(0, W)$, $Q(1, W)$ using the initial fit. Qinit\$type is method for estimating Q
g	treatment mechanism estimate. A list with three items: g\$g1W contains estimates of $P(A = 1 W)$ for each observation, g\$coef the coefficients for the model for g when glm used, g\$type estimation procedure
g.AV	estimate for $h(A, V)$ or $h(A, T)$. A list with three items: g.AV\$g1W an $n \times 2$ matrix containing values of $P(A = 0 V, T)$, $P(A = 1 V, T)$ for each observation, g.AV\$coef the coefficients for the model for g when glm used, g.AV\$type estimation procedure
g_Delta	missingness mechanism estimate. A list with three items: g_Delta\$g1W an $n \times 2$ matrix containing values of $P(\Delta = 1 A, V, W, T)$ for each observation, g_Delta\$coef the coefficients for the model for g when glm used, g_Delta\$type estimation procedure

Author(s)

Susan Gruber <sgruber@cal.berkeley.edu>, in collaboration with Mark van der Laan.

References

1. Gruber, S. and van der Laan, M.J. (2012), tmle: An R Package for Targeted Maximum Likelihood Estimation. *Journal of Statistical Software*, 51(13), 1-35. <http://www.jstatsoft.org/v51/i13/>
2. Rosenblum, M. and van der Laan, M.J. (2010), Targeted Maximum Likelihood Estimation of the Parameter of a Marginal Structural Model. *The International Journal of Biostatistics*, 6(2), 2010.

See Also

[summary.tmleMSM](#), [estimateQ](#), [estimateG](#), [calcSigma](#), [tmle](#)

Examples

```
library(tmle)
# Example 1. Estimating MSM parameter with correctly specified regression formulas
# MSM:  $\psi_0 + \psi_1*A + \psi_2*V + \psi_3*A*V$  (saturated)
# true parameter value:  $\psi = (0, 1, -2, 0.5)$ 
# generate data
set.seed(100)
n <- 1000
W <- matrix(rnorm(n*3), ncol = 3)
colnames(W) <- c("W1", "W2", "W3")
V <- rbinom(n, 1, 0.5)
A <- rbinom(n, 1, 0.5)
Y <- rbinom(n, 1, plogis(A - 2*V + 0.5*A*V))
result.ex1 <- tmleMSM(Y, A, W, V, MSM = "A*V", Qform = Y~., gform = A~1,
  hAVform = A~1, family = "binomial")
print(result.ex1)

# Example 2. Repeated measures data, two observations per id
# (e.g., crossover study design)
# MSM:  $\psi_0 + \psi_1*A + \psi_2*V + \psi_3*V^2 + \psi_4*T$ 
# true parameter value:  $\psi = (-2, 1, 0, -2, 0)$ 
# generate data in wide format (id, W1, Y(t), W2(t), V(t), A(t))
set.seed(100)
n <- 500
id <- rep(1:n)
W1 <- rbinom(n, 1, 0.5)
W2.1 <- rnorm(n)
W2.2 <- rnorm(n)
V.1 <- rnorm(n)
V.2 <- rnorm(n)
A.1 <- rbinom(n, 1, plogis(0.5 + 0.3 * W2.1))
A.2 <- 1-A.1
Y.1 <- -2 + A.1 - 2*V.1^2 + W2.1 + rnorm(n)
Y.2 <- -2 + A.2 - 2*V.2^2 + W2.2 + rnorm(n)
d <- data.frame(id, W1, W2=W2.1, W2.2, V=V.1, V.2, A=A.1, A.2, Y=Y.1, Y.2)
```

```

# change dataset from wide to long format
  longd <- reshape(d,
    varying = cbind(c(3, 5, 7, 9), c(4, 6, 8, 10)),
    idvar = "id",
    direction = "long",
    timevar = "T",
    new.row.names = NULL,
    sep = "")
# misspecified model for initial Q, partial misspecification for g
result.ex2 <- tmleMSM(Y = longd$Y, A = longd$A, W = longd[,c("W1", "W2")], V = longd$V,
  T = longd$T, MSM = "A + V + I(V^2) + T", Qform = Y ~ A + V, gform = A ~ W2, id = longd$id)
print(result.ex2)

# Example 3: Introduce 20% missingness in example 2 data
Delta <- rbinom(nrow(longd), 1, 0.8)
result.ex3 <- tmleMSM(Y = longd$Y, A = longd$A, W = longd[,c("W1", "W2")], V = longd$V, T=longd$T,
  Delta = Delta, MSM = "A + V + I(V^2) + T", Qform = Y ~ A + V, gform = A ~ W2,
  g.Deltaform = Delta~ 1, id=longd$id, verbose = TRUE)
print(result.ex3)

```

tmleNews

Show the NEWS file (tmleNews)

Description

Shows recent changes and bug fixes documented in the tmle package NEWS file.

Usage

```
tmleNews(...)
```

Arguments

... additional arguments passed to RShowDoc

Value

NONE

Author(s)

Susan Gruber

See Also

[tmle](#), [tmleMSM](#)

Index

*Topic **datasets**

fev, 8

calcParameters, 3, 6, 7, 15

calcSigma, 4, 4, 6, 7, 15, 19

estimateG, 4, 5, 5, 7, 15, 19

estimateQ, 4–6, 6, 15, 19

fev, 8

methods, 9, 11

print.summary.tmle(summary.tmle), 9

print.summary.tmleMSM
(summary.tmleMSM), 11

print.tmle(summary.tmle), 9

print.tmleMSM(summary.tmleMSM), 11

SL.glm.interaction, 8

summary.tmle, 9, 15

summary.tmleMSM, 11, 19

tmle, 3–7, 10, 12, 19, 20

tmle-package, 2

tmleMSM, 3–7, 12, 15, 16, 20

tmleNews, 20