Package 'multipred'

July 10, 2020

Type Package

Title Calculates measures of accuracy for risk predictors of multiple outcomes.

Version 0.2

Date 2020-07-10

Author Frank Dudbridge

Maintainer Frank Dudbridge <frank.dudbridge@leicester.ac.uk>

Description Accuracy can be evaluated in four senses: outcome-wise, joint, and panel-wise (weak sense and strong sense). For convenience the weak panel-wise sense is also called ``screening'', and the strong panel-wise sense simply ``panel-wise''. In each sense, accuracy can be measured empirically, within data sets given as input, or theoretically, given parameters of an underlying multivariate liability threshold model.

License GPL-3

LazyData TRUE

RoxygenNote 7.0.2

Suggests tmvtnorm

RemoteType github

RemoteHost api.github.com

RemoteRepo multipred

RemoteUsername DudbridgeLab

RemoteRef panelwise

RemoteSha 298edb0c99c82e992dbd392191ac60465aec97e8

GithubRepo multipred

GithubUsername DudbridgeLab

GithubRef panelwise

GithubSHA1 298edb0c99c82e992dbd392191ac60465aec97e8

NeedsCompilation no

Depends R (>= 3.5.0)

R topics documented:

`analyticJoint` *Analytic joint measures of predictive accuracy*

Description

Analytic calculation of joint sensitivity, specificity, positive and negative predictive value, concordance and relative utility, under a multivariate liability threshold model.

Usage

```
analyticJoint(VL, VX, VLX = NULL, thresh = NULL, prev, nsample = NULL)
```

Arguments

`VL`	Variance-covariance matrix of liability. Must have 1 on diagonal.
`VX`	Variance-covariance matrix of predictors.
`VLX`	Cross-covariance matrix between liabilities and predictors. Entry on row i, column j, is covariance between liability i and predictor j. Diagonal entries are the liability variances explained for each trait.
`thresh`	Vector of risk thresholds for predicting an event. If NULL, which is the default, concordance is the only measure that can be calculated.
`prev`	Vector of prevalences, ie population risks, for each trait.
`nsample`	Number of random pairs of samples drawn when estimating concordance. If NULL, which is the default, concordance is not calculated. If 0, all possible pairs are drawn from the data.

Details

Joint measures consider the prediction of all outcomes occuring in an individual. For example, joint sensitivity is the probability that, given an individual in which all outcomes did occur, the predicted risk exceeds the threshold for all outcomes. Joint specificity is the probability that, given an individual in which at least one outcome did not occur, the predicted risk is lower than the threshold for at least one outcome.

Joint concordance is the probability that, given one individual in which all outcomes did occur, and another in which at least one outcome did not occur, the minimum predicted risk over all outcomes is greater in the former individual. It is calculated by randomly simulating `nsample` such pairs of individuals from the specified model.

Value

A list with the following components

`sens` Sensitivity

`spec` Specificity

`PPV` Positive predictive value

`NPV` Negative predictive value

`C` Concordance

`RU` Relative utility

Examples

```
# results will vary due to random sampling in computing multvariate integrals
attach(PRSdata)
analyticJoint(VL,VX,VX,thresh=prevalence,prev=prevalence,nsample=1e5)

# $sens
# [1] 0.04205708

# $spec
# [1] 0.9958742

# $PPV
# [1] 6.884139e-09

# $NPV
# [1] 1

# $C
# [1] 0.76669

# $RU
# [1] 0.03931163
```

---

analyticOutcomeWise

Analytic outcome-wise measures of predictive accuracy

Description

Analytic calculation of outcome-wise sensitivity, specificity, positive and negative predictive value, concordance and relative utility, under a multivariate liability threshold model.

Usage

```
analyticOutcomeWise(VL, VX, VLX = NULL, thresh = NULL, weight = NULL, prev)
```

Arguments

VL	Variance-covariance matrix of liability. Must have 1 on diagonal.
VX	Variance-covariance matrix of predictors.
VLX	Cross-covariance matrix between liabilities and predictors. Entry on row i, column j, is covariance between liability i and predictor j. Diagonal entries are the liability variances explained for each trait.
thresh	Vector of risk thresholds for predicting an event. If NULL, which is the default, concordance is the only measure that can be calculated.
weight	Vector of weights.
prev	Vector of prevalences, ie population risks, for each trait.

Details

Outcome-wise measures consider the prediction of individual outcomes summed over individuals. When weight is a vector of 1's (default), outcome-wise measures correspond to classical univariate measures with the x matrix vectorised into a column vector. More generally, weight allows different outcomes to contribute more or less to the calculations.

Outcome-wise sensitivity, specificity and concordance are weighted sums of the univariate measures, where the weights depend on prev.

Value

A list with the following components

sens Sensitivity

spec Specificity

PPV Positive predictive value

NPV Negative predictive value

C Concordance

RU Relative utility

Examples

```
attach(PRSdata)
analyticOutcomeWise(VL,VX,VX,thresh=prevalence,prev=prevalence)

# $sens
# [1] 0.6243863

# $spec
# [1] 0.6132883

# $PPV
# [1] 0.04641913

# $NPV
# [1] 0.9818697

# $C
# [1] 0.6533142
```

```
# $RU
# [1]  0.2376747
```

`analyticPanelWise` *Analytic panel-wise measures of predictive accuracy*

Description

Analytic calculation of panel-wise sensitivity, specificity, positive and negative predictive value, concordance and relative utility, under a multivariate liability threshold model.

Usage

```
analyticPanelWise(VL, VX, VLX = NULL, thresh = NULL, prev, nsample = NULL)
```

Arguments

`VL`	Variance-covariance matrix of liability. Must have 1 on diagonal.
`VX`	Variance-covariance matrix of predictors.
`VLX`	Cross-covariance matrix between liabilities and predictors. Entry on row i, column j, is covariance between liability i and predictor j. Diagonal entries are the liability variances explained for each trait.
`thresh`	Vector of risk thresholds for predicting an event. If NULL, which is the default, concordance is the only measure that can be calculated.
`prev`	Vector of prevalences, ie population risks, for each trait.
`nsample`	Number of random pairs of samples drawn when estimating concordance. If NULL, which is the default, concordance is not calculated. If 0, all possible pairs are drawn from the data.

Details

Panel-wise measures consider the prediction of at least one outcome to occur. At least one outcome that did occur must be predicted to occur. For example, panel-wise sensitivity is the probability that, for an individual in which at least one outcome did occur, the predicted risk exceeds the threshold for at least one of the outcomes that did occur. Panel-wise specificity is the probability that, for an individual in which at least one outcome did not occur, the predicted risk is lower than the threshold for all the outcomes that did not occur.

Panel-wise concordance is the probability that given one individual in which at least one outcome did occur, and another in which at least one did not occur, the maximum predicted risk over all outcomes that occurred in the former is higher than the maximum over all outcomes that did not occur in the latter. Note that under this definition an individual can be either concordant or discordant with itself. Concordance is calculated by randomly simulating `nsample` such pairs of individuals from the specified model.

Value

A list with the following components

`sens` Sensitivity

`spec` Specificity

`PPV` Positive predictive value

`NPV` Negative predictive value

`C` Concordance

`RU` Relative utility

Examples

```
# results will vary due to random sampling in computing multivariate integrals
attach(PRSdata)
analyticPanelWise(VL,VX,VX,thresh=prevalence,prev=prevalence,nsample=1e5)

# $sens
# [1] 0.6463497

# $spec
# [1] 0.0708455

# $PPV
# [1] 0.1081343

# $NPV
# [1] 0.9371735

# $C
# [1] 0.49142

# $RU
# [1] -0.31006
```

`analyticScreening` *Analytic screening measures of predictive accuracy*

Description

Analytic calculation of screening sensitivity, specificity, positive and negative predictive value, concordance and relative utility, under a multivariate liability threshold model.

Usage

```
analyticScreening(VL, VX, VLX = NULL, thresh = NULL, prev, nsample = NULL)
```

Arguments

VL	Variance-covariance matrix of liability. Must have 1 on diagonal.
VX	Variance-covariance matrix of predictors.
VLX	Cross-covariance matrix between liabilities and predictors. Entry on row i, column j, is covariance between liability i and predictor j. Diagonal entries are the liability variances explained for each trait.
thresh	Vector of risk thresholds for predicting an event. If NULL, which is the default, concordance is the only measure that can be calculated.
prev	Vector of prevalences, ie population risks, for each trait.
nsample	Number of random samples drawn when estimating concordance. If NULL, which is the default, concordance is not calculated.

Details

Screening measures consider the prediction of at least one outcome to occur, without regard to whether the correct outcomes are predicted. For example, screening sensitivity is the probability that, for an individual in which at least one outcome did occur, the predicted risk exceeds the threshold for at least one outcome (but not necessary the ones that did occur). Screening specificity is the probability that, for an individual in which no outcomes did occur, the predicted risk is lower than the threshold for all outcomes.

Screening concordance is the probability that given one individual in which at least one outcome did occur, and another in which no outcomes did occur, the maximum predicted risk over all outcomes is higher in the former individual. It is calculated by randomly simulating nsample such pairs of individuals from the specified model.

Value

A list with the following components

sens Sensitivity

spec Specificity

PPV Positive predictive value

NPV Negative predictive value

C Concordance

RU Relative utility

Examples

```
# results will vary due to random sampling in computing multvariate integrals
attach(PRSdata)
analyticScreening(VL,VX,VX,thresh=prevalence,prev=prevalence,nsample=1e5)

# $sens
# [1] 0.9591925

# $spec
# [1] 0.06055228

# $PPV
# [1] 0.1604819
```

```
# $NPV
# [1] 0.8879618

# $C
# [1] 0.59799

# $RU
# [1] -0.04479532
```

`joint` *Joint measures of predictive accuracy*

Description

Calculates joint sensitivity, specificity, positive and negative predictive value, concordance and relative utility for a vector of predictors.

Usage

```
joint(x, y, thresh = NULL, prev = NULL, condprev = NULL, nsample = NULL)
```

Arguments

`x`	Matrix of predicted risks. Each row corresponds to an individual, each column to an outcome. Each entry should be a risk between 0 and 1.
`y`	Matrix of outcomes. Each row corresponds to an individual, each column to an outcome. Must contain binary outcomes coded as 0 and 1.
`thresh`	Vector of risk thresholds. For each row of x, each outcome is predicted to occur for which the risk exceeds the corresponding element of thresh. These predictions are then compared to the elements of y. If NULL, which is the default, concordance is the only measure that can be calculated.
`prev`	Probability of all events occurring, required for calculating relative utility. If NULL, which is the default, then `prev` is estimated from the `y` matrix, ignoring ascertainment.
`condprev`	Probability of all events occuring, conditional on the risk predictions being equal to `thresh`. If NULL, which is the default, then `prev` is set to the product of the elements of `thresh`. This working definition is exact when predictions and outcomes both are jointly independent.
`nsample`	Number of random pairs of samples drawn when estimating concordance. If NULL, which is the default, concordance is not calculated. If 0, all possible pairs are drawn from the data.

Details

Joint measures consider the prediction of all outcomes occuring in an individual. For example, joint sensitivity is the probability that, given an individual in which all outcomes did occur, the predicted risk exceeds the threshold for all outcomes. Joint specificity is the probability that, given an individual in which at least one outcome did not occur, the predicted risk is lower than the threshold for at least one outcome.

Joint concordance is the probability that, given one individual in which all outcomes did occur, and another in which at least one outcome did not occur, the minimum predicted risk over all outcomes is greater in the former individual. It is calculated by randomly drawing such pairs of individuals from `y`. If `nsample` is zero, all such pairs are drawn from `y`; this might be time-consuming. Therefore the default is not to calculate concordance. However, a good estimate of concordance can be obtained from a limited number of random samples `nsample`.

`prev` and `condprev` are only required to calculate relative utility, and can be omitted otherwise.

Value

A list with the following components

`sens` Sensitivity

`spec` Specificity

`PPV` Positive predictive value

`NPV` Negative predictive value

`C` Concordance

`RU` Relative utility

Examples

```
attach(PRSdata)
joint(risk[,1:2],disease[,1:2],thresh=prevalence[1:2],nsample=1e5)

# $sens
# [1] 0.4041096

# $spec
# [1] 0.7653745

# $PPV
# [1] 0.02488402

# $NPV
# [1] 0.9885961

# $C
# [1] 0.63282

# $RU
# [1] 0.3292956
```

`multipred` *multipred*

Description

Package for calculating measures of accuracy for risk predictors of multiple outcomes.

Details

Accuracy can be evaluated in four senses: outcome-wise, joint, and panel-wise (weak sense and strong sense). For convenience the weak panel-wise sense is also called "screening", and the strong panel-wise sense simply "panel-wise". In each sense, accuracy can be measured empirically, within data sets given as input, or theoretically, given parameters of an underlying multivariate liability threshold model.

Throughout the documentation, an "outcome" means one of several binary variables observed in an individual, and an outcome "occurs" when the variable has the positive state.

Outcome-wise measures calculate standard univariate measures of accuracy over all outcomes and individuals.

Joint measures consider the prediction of all outcomes occuring simultaneously within an individual.

Screening measures consider the prediction of at least one outcome occuring within an individual. It is not necessary that the predicted outcomes are the same ones that actually occur.

Panel-wise measures consider the prediction of at least one outcome occuring within an individual. There must be at least one predicted outcome that actually occurs.

Functions

`outcomeWise`

`joint`

`screening`

`panelWise`

`analyticOutcomeWise`

`analyticJoint`

`analyticScreening`

`analyticPanelWise`

`outcomeWise` *Outcome-wise measures of predictive accuracy*

Description

Calculates outcome-wise sensitivity, specificity, positive and negative predictive value, concordance and relative utility for a vector of predictors.

Usage

```
outcomeWise(x, y, thresh = NULL, weight = NULL, prev = NULL)
```

Arguments

x	Matrix of predicted risks. Each row corresponds to an individual, each column to an outcome. Each entry should be a risk between 0 and 1.
y	Matrix of outcomes. Each row corresponds to an individual, each column to an outcome. Must contain binary outcomes coded as 0 and 1.
thresh	Vector of risk thresholds. For each row of x, each outcome is predicted to occur for which the risk exceeds the corresponding element of thresh. These predictions are then compared to the elements of y. If NULL, which is the default, concordance is the only measure that can be calculated.
weight	Vector of weights. Defaults to a vector of 1's.
prev	Vector of prevalences, ie population risks, for each trait. Defaults to NULL, in which case prevalences are estimated in the data, ignoring ascertainment.

Details

Outcome-wise measures consider the prediction of individual outcomes summed over individuals. When `weight` is a vector of 1's (default), outcome-wise measures correspond to classical univariate measures with the `x` matrix vectorised into a column vector. More generally, `weight` allows different outcomes to contribute more or less to the calculations.

Outcome-wise sensitivity, specificity and concordance are weighted sums of the univariate measures, where the weights depend on `prev`. By default, `prev` is estimated from the outcome rates in `y`, but external estimates of population risk may be used instead.

Value

A list with the following components

`sens` Sensitivity

`spec` Specificity

`PPV` Positive predictive value

`NPV` Negative predictive value

`C` Concordance

`RU` Relative utility

Examples

```
attach(PRSdata)
outcomeWise(risk,disease,thresh=prevalence)

# $sens
# [1] 0.6017748

# $spec
# [1] 0.6129354

# $PPV
# [1] 0.04595316

# $NPV
# [1] 0.9802688
```

```
# $C
# [1] 0.6442582

# $RU
# [1] 0.2251043
```

`panelWise` *Panel-wise measures of predictive accuracy*

Description

Calculates panel-wise sensitivity, specificity, positive and negative predictive value, concordance and relative utility for a vector of predictors.

Usage

```
panelWise(
  x,
  y,
  thresh = NULL,
  prev0 = NULL,
  prev1 = NULL,
  condprev0 = NULL,
  condprev1 = NULL,
  nsample = NULL
)
```

Arguments

`x`	Matrix of predicted risks. Each row corresponds to an individual, each column to an outcome. Each entry should be a risk between 0 and 1.
`y`	Matrix of outcomes. Each row corresponds to an individual, each column to an outcome. Must contain binary outcomes coded as 0 and 1.
`thresh`	Vector of risk thresholds. For each row of x, each outcome is predicted to occur for which the risk exceeds the corresponding element of thresh. These predictions are then compared to the elements of y. If NULL, which is the default, concordance is the only measure that can be calculated.
`prev0`	Probability of at least one non-event, required for calculating relative utility. If NULL, which is the default, then `prev0` is estimated from the `y` matrix, ignoring ascertainment.
`prev1`	Probability of at least one event, required for calculating relative utility. If NULL, which is the default, then `prev1` is estimated from the `y` matrix, ignoring ascertainment.
`condprev0`	Probability of at least one non-event, conditional on the risk predictions being equal to `thresh`. If NULL, which is the default, then `prev` is set to 1 - the product of the elements of `thresh`. This working definition is exact when predictions and outcomes both are jointly independent.

`condprev1` Probability of at least one event, conditional on the risk predictions being equal to `thresh`. If NULL, which is the default, then `prev` is set to 1 - the product of the elements of (1-`thresh`). This working definition is exact when predictions and outcomes both are jointly independent.

`nsample` Number of random pairs of samples drawn when estimating concordance. If NULL, which is the default, concordance is not calculated. If 0, all possible pairs are drawn from the data.

Details

Panel-wise measures consider the prediction of at least one outcome to occur. At least one outcome that did occur must be predicted to occur. For example, panel-wise sensitivity is the probability that, for an individual in which at least one outcome did occur, the predicted risk exceeds the threshold for at least one of the outcomes that did occur. Panel-wise specificity is the probability that, for an individual in which at least one outcome did not occur, the predicted risk is lower than the threshold for all the outcomes that did not occur.

Panel-wise concordance is the probability that given one individual in which at least one outcome did occur, and another in which at least one did not occur, the maximum predicted risk over all outcomes that occurred in the former is higher than the maximum over all outcomes that did not occur in the latter. Note that under this definition an individual can be either concordant or discordant with itself. Concordance is calculated by randomly drawing such pairs of individuals from `y`. If `nsample` is zero, all such pairs are drawn from `y`; this might be time-consuming. Therefore the default is not to calculate condcordance. However, a good estimate of concordance can be obtained from a limited number of random samples `nsample`.

`prev0`, `prev1`, `condprev0` and `condprev1` are only required to calculate relative utility, and can be omitted otherwise.

Value

A list with the following components

`sens` Sensitivity

`spec` Specificity

`PPV` Positive predictive value

`NPV` Negative predictive value

`C` Concordance

`RU` Relative utility

Examples

```
attach(PRSdata)
panelWise(risk,disease,thresh=prevalence,nsample=1e5)

# $sens
# [1] 0.6266996

# $spec
# [1] 0.0701

# $PPV
# [1] 0.1073925
```

```
# $NPV
# [1] 0.9311696

# $C
# [1] 0.47999

# $RU
# [1] -5.120519
```

`PRSdata` *Polygenic risk scores*

Description

Simulated liabilities and polygenic risk scores for six complex diseases, with corresponding disease outcomes and risk predictions.

Usage

```
PRSdata
```

Format

A list with 7 elements. Each element is a matrix with 10,000 rows and 6 columns corresponding to simulated data on 10,000 individuals for 6 diseases. Diseases are modelled on type-2 diabetes, coronary artery disease, Crohn's disease, ulcerative colitis, schizophrenia and rheumatoid arthritis.

`PRS` Polygenic risk scores

`liability` Liabilities

`risk` Disease risks corresponding to `PRS`

`disease` Disease outcomes corresponding to `liability`

`VL` Variance-covariance matrix of `liability`

`VX` Variance-covariance matrix of `PRS`

`h2` Disease heritabilities

`prevalence` Disease prevalences

screening *Screening measures of predictive accuracy*

Description

Calculates screening sensitivity, specificity, positive and negative predictive value, concordance and relative utility for a vector of predictors.

Usage

```
screening(x, y, thresh = NULL, prev = NULL, condprev = NULL, nsample = NULL)
```

Arguments

x	Matrix of predicted risks. Each row corresponds to an individual, each column to an outcome. Each entry should be a risk between 0 and 1.
y	Matrix of outcomes. Each row corresponds to an individual, each column to an outcome. Must contain binary outcomes coded as 0 and 1.
thresh	Vector of risk thresholds. For each row of x, each outcome is predicted to occur for which the risk exceeds the corresponding element of thresh. These predictions are then compared to the elements of y. If NULL, which is the default, concordance is the only measure that can be calculated.
prev	Probability of at least one event, required for calculating relative utility. If NULL, which is the default, then prev is estimated from the y matrix, ignoring ascertainment.
condprev	Probability of at least one events, conditional on the risk predictions being equal to thresh. If NULL, which is the default, then prev is set to 1- the product of the elements of (1-thresh). This working definition is exact when predictions and outcomes both are jointly independent.
nsample	Number of random pairs of samples drawn when estimating concordance. If NULL, which is the default, concordance is not calculated. If 0, all possible pairs are drawn from the data.

Details

Screening measures consider the prediction of at least one outcome to occur, without regard to whether the correct outcomes are predicted. For example, screening sensitivity is the probability that, for an individual in which at least one outcome did occur, the predicted risk exceeds the threshold for at least one outcome (but not necessary the ones that did occur). Screening specificity is the probability that, for an individual in which no outcomes did occur, the predicted risk is lower than the threshold for all outcomes.

Screening concordance is the probability that given one individual in which at least one outcome did occur, and another in which no outcomes did occur, the maximum predicted risk over all outcomes is higher in the former individual. It is calculated by randomly drawing such pairs of individuals from y. If nsample is zero, all such pairs are drawn from y; this might be time-consuming. Therefore the default is not to calculate concordance. However, a good estimate can be obtained from a limited number of random samples nsample.

prev and condprev are only required to calculate relative utility, and can be omitted otherwise.

Value

A list with the following components

`sens` Sensitivity

`spec` Specificity

`PPV` Positive predictive value

`NPV` Negative predictive value

`C` Concordance

`RU` Relative utility

Examples

```
attach(PRSdata)
screening(risk,disease,thresh=prevalence,nsample=1e5)

# $sens
# [1] 0.9653894

# $spec
# [1] 0.05989024

# $PPV
# [1] 0.1654311

# $NPV
# [1] 0.8996416

# $C
# [1] 0.59156

# $RU
# [1] -0.009099365
```

Index