Instructions on Running the Code for the Proposed Penalized Empirical Likelihood Approach

We have proposed three approaches to combine information from high-dimensional subject-level data and external aggregated data:

- PEL: the proposed penalized empirical likelihood approach

- PEL*: the proposed penalized empirical likelihood approach that accounts for uncertainty in the external aggregated information

- PEL$_{\boldsymbol{\alpha}}$: the proposed penalized empirical likelihood approach that accounts for heterogeneity in the predictor distribution

1 The Proposed PEL and PEL* Approaches

1.1 The *pel* Function

The *pel* function is used to calculate the proposed PEL and PEL* estimators.

1.1.1 Input

- $\boldsymbol{x}$: a $n \times (p + 1)$ matrix of predictors

- y: a $n \times 1$ vector of binary response

- info: the external aggregated information

- lambda: the tuning parameter for the adaptive lasso penalty

- v: the exponent in the adaptive weight

- b.ini: the initial value for parameters, which is used to construct the adaptive weight

- index: the index of predictors whose summary statistics are provided

1.1.2 Output

- b.new: the estimates of the parameters in the logistic regression

- xi.new: the estimates of the Lagrange multipliers

1.2 The *bic* Function

The *bic* function is used to calculate the BIC type criteria value for the PEL or PEL* estimator.

1.2.1 Input

- $\boldsymbol{x}$: a $n \times (p+1)$ matrix of predictors

- y: a $n \times 1$ vector of binary response

- b: the estimates of the parameters

- xi: the estimates of the Lagrange multipliers

- info: the external aggregated information

- index: the index of predictors whose summary statistics are provided

1.2.2 Output

- the BIC type criteria value

1.3 The *get.cov* Function

The *get.cov* function is used to calculate the standard error estimates for the nonzero parameters using the proposed PEL and PEL* approaches.

1.3.1 Input

- kappa: n/m, where n is the sample size of the subject-level data and m is the sample size of external aggregate data

- setA: the index of nonzero parameters

- $\boldsymbol{x}$: a $n \times (p+1)$ matrix of predictors

- b.pel: the estimates of the parameters using the PEL approach

- b.star: the estimates of the parameters using the PEL* approach

- info: the external aggregated information used by the PEL approach

- info.star: the external aggregated information used by the PEL* approach

- index: the index of predictors whose summary statistics are provided

- b.ini: the initial value for parameters, which is used to construct the adaptive weight

- lambda1: the tuning parameter for the adaptive lasso penalty using the PEL approach

- lambda2: the tuning parameter for the adaptive lasso penalty using the PEL* approach

- v1: the exponent in the adaptive weight using the PEL approach

- v2: the exponent in the adaptive weight using the PEL* approach

1.3.2 Output

- se.pel: the standard error estimates for the nonzero parameters using the PEL approach

- se.star: the standard error estimates for the nonzero parameters using the PEL* approach

2 The Proposed PEL$_\alpha$ Approach

2.1 The *pel.ex* Function

The *pel.ex* function is used to calculate the proposed PEL$_{\boldsymbol{\alpha}}$ estimator.

2.1.1 Input

- $\boldsymbol{x}$: a $n \times (p+1)$ matrix of predictors

- y: a $n \times 1$ vector of binary response

- info: the external aggregated information

- tau: the tuning parameter for the adaptive lasso penalty

- v: the exponent in the adaptive weight

- setB: the index of the predictors in the density ratio model

- b.ini: the initial value for parameters, which is used to construct the adaptive weight

- index: the index of predictors whose summary statistics are provided

2.1.2 Output

- b.new: the estimates of the parameters in the logistic regression

- a.new: the estimates of the parameters in the density ratio model

- eta.new: the estimates of the Lagrange multipliers

2.2 The *bicex* Function

The *bicex* function is used to calculate the BIC type criteria value for the proposed $\text{PEL}_{\boldsymbol{\alpha}}$ estimator.

2.2.1 Input

- $\boldsymbol{x}$: a $n \times (p+1)$ matrix of predictors

- y: a $n \times 1$ vector of binary response

- a: the estimates of the parameters in the density ratio model

- b: the estimates of the parameters in the logistic regression

- eta: the estimates of the Lagrange multipliers

- info: the external aggregated information

- setB: the index of the predictors in the density ratio model

- index: the index of predictors whose summary statistics are provided

2.2.2 Output

- the BIC type criteria value

2.3   The *get.cov.ex* Function

The *get.cov.ex* function is used to calculate the standard error estimates for the nonzero parameters using the proposed $\text{PEL}_{\boldsymbol{\alpha}}$ approach.

2.3.1   Input

- setA: the index of nonzero parameters in the logistic regression

- setB: the index of the predictors in the density ratio model

- $\boldsymbol{x}$: a $n \times (p+1)$ matrix of predictors

- b.ini: the initial value for parameters, which is used to construct the adaptive weight

- b.ex: the estimates of the parameters in the logistic regression using the $\text{PEL}_{\boldsymbol{\alpha}}$ approach

- a.ex: the estimates of the parameters in the density ratio model using the $\text{PEL}_{\boldsymbol{\alpha}}$ approach

- info: the external aggregated information

- index: the index of predictors whose summary statistics are provided

- tau: the tuning parameter for the adaptive lasso penalty using the $\text{PEL}_{\boldsymbol{\alpha}}$ approach

- v: the exponent in the adaptive weight using the $\text{PEL}_{\boldsymbol{\alpha}}$ approach

2.3.2   Output

- the standard error estimates for the nonzero parameters using the $\text{PEL}_{\boldsymbol{\alpha}}$ approach