Package 'fabPrediction'

March 26, 2024

Title Compute FAB (Frequentist and Bayes) Conformal Prediction Intervals

Version 1.0.4

Description Computes and plots prediction intervals for numerical data or prediction sets for categorical data using prior information. Empirical Bayes procedures to estimate the prior information from multi-group data are included. See, e.g.,Bersson and Hoff (2022) <arXiv:2204.08122> ``Optimal Conformal Prediction for Small Areas".

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.3

Imports sae (>= 1.3), parallel, stats, graphics

Suggests knitr, devtools

NeedsCompilation yes

ByteCompile yes

VignetteBuilder knitr

URL https://github.com/betsybersson/fabPrediction

BugReports https://github.com/betsybersson/fabPrediction/issues

Author Elizabeth Bersson [aut, cre, cph]

Maintainer Elizabeth Bersson <elizabeth.bersson@duke.edu>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2024-03-26 09:00:05 UTC

R topics documented:

`bayesMultinomialPrediction` *Obtain a Bayesian prediction interval for categorical data*

Description

This function computes the Bayesian prediction set for a multinomial conjugate family.

Usage

```
bayesMultinomialPrediction(
  Y,
  alpha = 0.15,
  gamma = rep(1, length(Y)),
  category_names = 1:length(Y)
)
```

Arguments

`Y`	Observed data vector of length K containing counts of observations from each of the K categories
`alpha`	Prediction mis-coverage rate
`gamma`	Dirichlet prior concentration for the K categories
`category_names`	Category names (optional)

Value

pred object

`bayesNormalPrediction` *Obtain a Bayesian prediction interval*

Description

This function computes a Bayesian prediction interval based on a normal model.

Usage

```
bayesNormalPrediction(Y, alpha = 0.15, mu = 0, tau2 = 1)
```

Arguments

`Y`	Observed data vector
`alpha`	Prediction error rate
`mu`	Prior expected mean of the population mean
`tau2`	Prior expected variance of the population mean

Value

pred object

`dtaPrediction` *Obtain a distance-to-average conformal prediction interval*

Description

This function computes a conformal prediction region under the distance-from-average non-conformity measure. That is, |a + bz*| <= |ci + di z^*| where i indexes training data.

Usage

```
dtaPrediction(Y, alpha = 0.15)
```

Arguments

`Y`	Observed data vector
`alpha`	Prediction error rate

Value

pred object

eye *Create Identity Matrix*

Description

This function returns an NxN identity matrix.

Usage

```
eye(N)
```

Arguments

N	dimension of square matrix

Value

NxN identity matrix

fabCategoricalPrediction *Obtain a FAB conformal prediction interval for categorical data*

Description

This function computes a FAB conformal prediction set as described in Bersson and Hoff 2023.

Usage

```
fabCategoricalPrediction(
  Y,
  alpha = 0.15,
  gamma = rep(1, length(Y)),
  category_names = 1:length(Y)
)
```

Arguments

Y	Observed data vector of length K containing counts of observations from each of the K categories
alpha	Prediction mis-coverage rate
gamma	Dirichlet prior concentration for the K categories
category_names	Category names (optional)

Value

pred object

`fabContinuousPrediction` *Obtain a FAB conformal prediction interval*

Description

This function computes a FAB conformal prediction region as described in Bersson and Hoff 2022.

Usage

```
fabContinuousPrediction(Y, alpha = 0.15, mu = 0, tau2 = 1)
```

Arguments

`Y`	Observed data vector
`alpha`	Prediction error rate
`mu`	Prior expected mean of the population mean
`tau2`	Prior expected variance of the population mean

Value

pred object

`fabPrediction` *fabPrediction: Compute FAB Conformal Prediction Intervals*

Description

A package for computing and plotting prediction intervals for numerical data or prediction sets for categorical data using prior information. Empirical Bayes procedures to estimate the prior information from multi-group data are included.

Author(s)

Elizabeth Bersson,
Maintainer: Elizabeth Bersson <elb75@duke.edu>

References

E. Bersson and P.D. Hoff. (2023) Frequentist Prediction Sets for Species Abundance using Indirect Information. Preprint.

E. Bersson and P.D. Hoff. (2023) Optimal Conformal Prediction for Small Areas. Journal of Survey Statistics and Methodology, forthcoming.

`fayHerriotEB` *Obtain empirical Bayesian estimates for group j*

Description

This function returns empirical Bayesian estimates for a specified group from the conjugate normal spatial Fay-Herriot model.

Usage

```
fayHerriotEB(j, Y, group, W = NA, X = NA)
```

Arguments

`j`	Obtain EB values for group in index j- numeric value in group
`Y`	Data vector
`group`	index vecter of the same lenght as Y
`W`	Non-standardized adjacency matrix
`X`	Group-level covariates

Value

empirical Bayesian estimates of population mean and it's variance

`initMoM` *Obtain inital guess of MLE of the marginal Dirichlet-multinomial likelihood*

Description

Method of moment matching to obtain an initial guess of the MLE, as in Minka (2000).

Usage

```
initMoM(D)
```

Arguments

`D`	matrix (JxK) of counts; each row is a sample from a MN distribution with K categories

Value

Hessian

normalPrediction *Obtain a pivot prediction interval*

Description

This function computes a prediction interval under assumed normality.

Usage

```
normalPrediction(Y, alpha = 0.15)
```

Arguments

Y	Observed data vector
alpha	Prediction error rate

Value

pred object

plot.pred *Plot a 'pred' object constructed for a categorical response*

Description

Plot a 'pred' object constructed for a categorical response

Usage

```
## S3 method for class 'pred'
plot(x, ...)
```

Arguments

x	pred object- a list classified as pred containing objects data and bound
...	additional parameters passed to the default plot method

Value

capability to plot pred object. More details: the command 'plot(obj)' plots the empirical densities of each category. Mass denoted in red indicates inclusion in the prediction set

pluginValues — *Obtain empirical Bayesian estimates for conjugate normal spatial Fay-Herriot model*

Description

This function returns plug-in values for a conjugate normal spatial Fay-Herriot model.

Usage

```
pluginValues(Y, group, W = NA, X = NA)
```

Arguments

Y	Data vector
group	Group membership of each entry in Y
W	Adjacency matrix
X	Group-level covariates

Value

plug-in values of spatial Fay-Herriot model

polyaGradient — *Obtain gradient of the marginal Dirichlet-multinomial likelihood*

Description

Obtain gradient of the marginal Dirichlet-multinomial likelihood

Usage

```
polyaGradient(D, gamma, Nj = rowSums(D), K = ncol(D))
```

Arguments

D	matrix (JxK) of counts; each row is a sample from a MN distribution with K categories
gamma	current value of prior concentration parameter
Nj	sample sizes of the J groups
K	number of categories

Value

gradient

polyaHessian *Obtain Hessian of the marginal Dirichlet-multinomial likelihood*

Description

Obtain Hessian of the marginal Dirichlet-multinomial likelihood

Usage

```
polyaHessian(D, gamma, Nj = rowSums(D), K = ncol(D))
```

Arguments

D	matrix (JxK) of counts; each row is a sample from a MN distribution with K categories
gamma	current value of prior concentration parameter
Nj	sample sizes of the J groups
K	number of categories

Value

Hessian

polyaMLE *Obtain MLE of marginal Dirichlet-multinomial likelihood*

Description

This function retuns the MLE of the prior concentration from a marginal Dirichlet-multinomial likelihood. Default method iterates a Newton-Raphson algorithm until convergence.

Usage

```
polyaMLE(
  D,
  init = NA,
  method = "Newton_Raphson",
  epsilon = 1e-04,
  print_progress = FALSE
)
```

Arguments

D	matrix (JxK) of counts; each row is a sample from a MN distribution with K categories
init	If NA, use method moment matching procedure to obtain good init values
method	"Newton_Raphson", "fixed_point", "separate", "precision_only"
epsilon	convergence diagnostic
print_progress	if TRUE, print progress to screen

Value

mle of prior concentration from marginal Dirichlet-multinomial likelihood

---

predictionInterval *Wrapper to obtain a prediction interval for continuous data*

Description

This function computes a prediction interval from a number of methods.

Usage

```
predictionInterval(Y, method = "FAB", alpha = 0.15, mu = 0, tau2 = 1)
```

Arguments

Y	Observed data vector
method	Choice of prediction method. Options include FAB, DTA, direct, Bayes.
alpha	Prediction error rate
mu	Prior expected mean of the population mean
tau2	Prior expected variance of the population mean

Value

pred object containing prediction interval bounds and interval coverage

Examples

```
# example data
data(radon)
y_county9 = radon$radon[radon$group==9]

fab.region = predictionInterval(y_county9,
  method = "FAB",
  alpha = .15,
  mu = 0.5,tau2 = 1)
```

```
fab.region$bounds
plot(fab.region)
```

---

predictionSet *Wrapper to obtain a prediction set for categorical data*

Description

This function computes a prediction set from a number of methods.

Usage

```
predictionSet(
  Y,
  method = "FAB",
  alpha = 0.15,
  gamma = rep(1, length(Y)),
  category_names = 1:length(Y)
)
```

Arguments

Y	Observed data vector
method	Choice of prediction method. Options include FAB, direct, Bayes.
alpha	Prediction mis-coverage rate
gamma	Dirichlet prior concentration for FAB/Bayes methods
category_names	Category names (optional)

Value

pred object containing prediction set and interval coverage

Examples

```
# obtain example categorical data
set.seed(1)
prob = rdirichlet(50:1)
y = rmultinom(1,15,prob)

fab.set = predictionSet(y,
  method = "FAB",
  gamma = c(50:1))
plot(fab.set)
```

radon *Minnesota Radon Data*

Description

Data from a national US EPA survey of household radon values. County index contained in group column.

Usage

```
data(radon)
```

Format

A matrix.

Source

ARM Data

References

US Environmental Protection Agency (1992) National residential radon survey: summary report. Washington, DC; DOI EPA402-R-92-011.

rdirichlet *Generate a random sample from a Dirichlet distribution*

Description

Generate a random sample from a Dirichlet distribution

Usage

```
rdirichlet(gamma)
```

Arguments

gamma	Prior concentration vector of length K

Value

a vector of length K that is a random sample from a Dirichlet distribution

row_standardize *Row standardize a matrix*

Description

Row standardize a matrix

Usage

```
row_standardize(W)
```

Arguments

W	matrix

Value

row-standardized matrix

W *Minnesota County Adjacency Matrix*

Description

Adjacency matrix for MN counties based on group index that matches radon data.

Usage

```
data(W)
```

Format

A matrix.

Index