Package 'slgf'

July 23, 2025

Type Package

Title Bayesian Model Selection with Suspected Latent Grouping Factors

Version 2.0.0

Date 2022-11-01

Description Implements the Bayesian model selection method with suspected latent
grouping factor methodology of Metzger and Franck (2020),
<doi:10.1080/00401706.2020.1739561>. SLGF detects latent
heteroscedasticity or group-based regression effects based on the levels of a
user-specified categorical predictor.

License GPL (>= 2)

Encoding UTF-8

Imports Rdpack, numDeriv, utils

RdMacros Rdpack

LazyData true

Depends R (>= 3.5.0)

Suggests knitr, captioner, formatR, rcrossref, rmarkdown

RoxygenNote 7.2.2

NeedsCompilation no

Author Thomas Metzger [aut, cre],
Christopher Franck [aut]

Maintainer Thomas Metzger <metzger.181@osu.edu>

Repository CRAN

Date/Publication 2022-11-21 22:30:02 UTC

Contents

---

bottles   *Bottles data on filling amount of six bottles over five time points.*

Description

Ott and Snee (1973) analyzed the filled weight (weight) of bottles filled by six machine heads (heads) over five time points (time).

Usage

```
data(bottles)
```

Format

A data frame with 30 rows and 3 variables:

weight the response, the weight of material filled into each bottle.

time the time point of each filling.

heads the head of the machine that fills the bottle.

References

Ott ER, Snee RD (1973). "Identifying useful differences in a multiple-head machine." *Journal of Qualitym Technology*, **5**(2), 2–13.

---

lymphoma   *Lymphoma data on genomic hybridization signal from six dogs with normal and tumor tissue samples taken.*

Description

Franck et. al. (2013) analyzes the genomic hybridization signal measured from normal and tumor tissue samples taken from six dogs.

Usage

```
data(lymphoma)
```

Format

A data frame with 6 rows and 2 variables:

V1 the signals from the normal tissue samples.

V2 the signals from the tumor tissue samples.

References

Franck CT, Nielsen DM, Osborne JA (2013). "A method for detecting hidden additivity in two-factor unreplicated experiments." *Computational Statistics & Data Analysis*, **67**(Supplement C), 95 - 104. ISSN 0167-9473, doi:10.1016/j.csda.2013.05.002.

---

maketall	*Converts a two-way layout into tall format with row and column index labels.*

Description

`maketall` Converts a two-way layout into tall format with row and column index labels.

Usage

```
maketall(data_matrix)
```

Arguments

`data_matrix`   an r by c data matrix.

Value

`maketall` returns a data frame containing the original observations, row labels, and column labels.

Examples

```
library(slgf)
data(lymphoma)
maketall(lymphoma)
```

| ms_slgf | *Bayesian Model Selection with Latent Group-Based Regression Effects and Heteroscedasticity* |

Description

`ms_slgf` Implements the model selection method proposed by (Metzger and Franck 2019).

Usage

```
ms_slgf(
  dataf,
  response,
  lgf_beta,
  min_levels_beta = 1,
  lgf_Sigma,
  min_levels_Sigma = 1,
  same_scheme = TRUE,
  usermodels,
  het = rep(0, length(usermodels)),
  prior = "flat",
  m0 = NULL
)
```

Arguments

`dataf`	A data frame containing a continuous response, at least one categorical predictor, and any other covariates of interest. This data frame should not contain column names with the character string `group`.
`response`	A character string indicating the column of `dataf` that contains the response.
`lgf_beta`	An optional character string indicating the column of 'dataf' that contains the suspected latent grouping factor (SLGF) for the regression effects.
`min_levels_beta`	
	A numeric value indicating the minimum number of levels of 'lgf_beta' that can comprise a group. Defaults to 1.
`lgf_Sigma`	An optional character string indicating the column of 'dataf' that contains the suspected latent grouping factor (SLGF) for the residual variances.
`min_levels_Sigma`	
	A numeric value indicating the minimum number of levels of 'lgf_Sigma' that can comprise a group. Defaults to 1.
`same_scheme`	A Boolean operator indicating whether the schemes for 'lgf_beta' and 'lgf_Sigma' must be the same.
`usermodels`	A list of length `M` where each element contains a string of *R* class `formula` or `character` indicating the models to consider. The term `group` should be used to replace the name of the SLGF in models with group-based regression effects.

het	A vector of 0s and 1s of length M. If the mth element of `het` is 0, then the mth model of `usermodels` is considered in a homoscedastic context only; if the mth element of `het` is 1, the mth model of `usermodels` is considered in both homoscedastic and heteroscedastic contexts.
prior	A character string `"flat"` or `"zs"` indicating whether to implement the flat or Zellner-Siow mixture g-prior on regression effects, respectively. Defaults to `"flat"`.
m0	An integer value indicating the minimum training sample size. Defaults to NULL. If no value is provided, the lowest value that leads to convergence for all considered posterior model probabilities will be used. If the value provided is too low for convergence, it will be increased automatically.

Value

`ms_slgf` returns a list of five elements if the flat prior is used, and six elements if the Zellner-Siow mixture g-prior is used:

1) models, an M by 7 matrix where columns contain the model selection results and information for each model, including:
- `Model`, the formula associated with each model;
- `Scheme.beta`, the grouping scheme associated with the fixed effects;
- `Scheme.Sigma`, the grouping scheme associated with the variances;
- `Log-Marginal`, the fractional log-marginal likelihood associated with each model;
- `FmodProb`, the fractional posterior probability associated with each model;
- `ModPrior`, the prior assigned to each model;
- `Cumulative`, the cumulative fractional posterior probability associated with a given model and the previous models;

2) class_probabilities, a vector containing cumulative posterior probabilities associated with each model class;

3) coefficients, MLEs for each model's regression effects;

4) variances, MLEs based on concentrated likelihood for each model's variance(s);

5) gs, MLEs based on concentrated likelihood for each model's g; only included if `prior="zs"`.

Author(s)

Thomas A. Metzger and Christopher T. Franck

References

Metzger TA, Franck CT (2019). "Detection of latent heteroscedasticity and group-based regression effects in linear models via Bayesian model selection." *arXiv e-prints*.

Examples

```
# Analyze the smell and textile data sets.

library(numDeriv)


data(smell)
```

```
out_smell <- ms_slgf(dataf = smell, response = "olf", het=c(1,1),
                       lgf_beta = "agecat", lgf_Sigma = "agecat",
                       same_scheme=TRUE, min_levels_beta=1, min_levels_Sigma=1,
                       usermodels = list("olf~agecat", "olf~group"), m0=4)
out_smell$models[1:5,]
out_smell$coefficients[[46]]
out_smell$variances[[46]]

# textile data set
data(textile)
out_textile <- ms_slgf(dataf = textile, response = "strength",
                       lgf_beta = "starch", lgf_Sigma = "starch",
                       same_scheme=FALSE, min_levels_beta=1, min_levels_Sigma=1,
                       usermodels = list("strength~film+starch", "strength~film*starch",
                                         "strength~film+group", "strength~film*group"),
                       het=c(1,1,1,1), prior="flat", m0=8)
out_textile$models[1:5,c(1,2,3,5)]
out_textile$class_probabilities
out_textile$coefficients[31]
out_textile$variances[31]
```

---

roadwear	*Roadwear data on four tires, each comprising three compounds, from a balanced incompleted block design.*

Description

Davies (1954) analyzes the wear on four tires, where each tire comprises three distinct compounds.

Usage

```
data(roadwear)
```

Format

A data frame with 12 rows and 3 variables:

abrasion the measurement of abrasion.

compound the compound from which each measurement was taken, either A, B, C, or D.

tire the tire from which each measurement was taken, either 1, 2, 3, or 4.

References

Davies OL (1954). *The Design and Analysis of Industrial Experiments*. Oliver and Boyd, London.

---

smell   *Smell data on olfactory function by age group*

Description

O'Brien and Heft (1995) studied the University of Pennsylvania Smell Identification Test (UPSIT). 180 subjects of different age groups were asked to describe 40 different odors. Olfactory index was quantified by the Freeman-Tukey modified arcsine transformation on the proportion of correctly identified odors. Subjects were divided into five age groups: group 1 if age 2 or younger; group 2 if between ages 26 and 40; group 3 if between ages 41 and 55; group 4 if between ages 56 and 70; and group 5 if older than 75.

Usage

```
data(smell)
```

Format

A data frame with 180 rows and 2 variables:

agecat age category, from 1 to 5.

olf olfactory function, measured as the Freeman-Tukey modified arcsine transformation on the proportion of correctly identified odors.

Source

SAS/STAT 15.2 User's Guide

References

OBrien RG, Heft MW (1995). "New Discrimination Indexes and Models for Studying Sensory Functioning in Aging." *Journal of Applied Statistics*, **22**, 9-27.

---

textile   *Textile data on breaking strength by starch type and chip thickness*

Description

Furry (1939) analyzes the breaking strength of a starch chip as a function of the chip's thickness (measured in 10^-4 inches) and the type of plant from which the starch was derived (corn, canna, or potato).

Usage

```
data(textile)
```

Format

A data frame with 49 rows and 3 variables:

strength the response, the breaking strength.

film the chip's film thickness, measured in 10^-4 inches.

starch the chip's starch component: canna, corn, or potato

References

Furry MS (1939). "Breaking strength, elongation and folding endurance of films of starches and gelatin used in sizing." *Technical Bulletin (United States Department of Agriculture)*, **674**, 1-36.

---

torque   *Locknut data on torque required to tighten a fixture by plating method*

Description

Meek and Ozgur (1991) analyzes the torque required to strengthen a fixture (bolt or mandrel) as a function of the fixture's plating method (cadmium and wax, heat treating, and phosphate and oil, denoted CW, HT, and PO, respectively).

Usage

```
data(torque)
```

Format

A data frame with 60 rows and 3 variables:

Torque the response, the torque required to tighten the fixture.

Fixture the type of fixture, bolt or mandrel.

Plating the plating treatment, CW, HT, or PO.

References

Meek GE, Ozgur CO (1991). "Torque Variation Analysis." *Journal of the Industrial Mathematics Society*, **41**, 1-16.

Index