Package ‘mvglmmRank’

January 8, 2023

Type Package

Title Multivariate Generalized Linear Mixed Models for Ranking Sports Teams

Version 1.2-4

Depends R (>= 3.2.0), Matrix

Imports numDeriv, methods, stats, utils, MASS

Date 2023-01-06


ByteCompile yes

LazyLoad yes

LazyData yes

License GPL-2

NeedsCompilation no

Author Andrew T. Karl [cre, aut] (<https://orcid.org/0000-0002-5933-8706>), Jennifer Broatch [aut]

Maintainer Andrew T. Karl <akarl@asu.edu>

Repository CRAN

Date/Publication 2023-01-08 00:20:02 UTC

R topics documented:

mvglmmRank-package ............................................. 2
binary_cre .................................................... 3
f2008 .......................................................... 4
f2009 .......................................................... 5
Description

Maximum likelihood estimates are obtained via an EM algorithm with either a first-order or a fully exponential Laplace approximation as documented by Broatch and Karl (2018), Karl, Yang, and Lohr (2014), and by Karl (2012). Karl and Zimmerman use this package to illustrate how the home field effect estimator from a mixed model can be biased under nonrandom scheduling.

Details

Package: mvglmmRank
Type: Package
Version: 1.2-4
Date: 2023-01-06
License: GPL-2

See the help pages for mvglmmRank and game.pred

Author(s)

Andrew T. Karl and Jennifer Broatch

Maintainer: Andrew T. Karl <akarl@asu.edu>
References


Examples

data(nfl2012)
mvglmRank(nfl2012,method="PB0",first.order=TRUE,verbose=TRUE,max.iter.EM=1)

result <- mvglmRank(nfl2012,method="PB0",first.order=TRUE,verbose=TRUE)
print(result)
game.pred(result,home="Denver Broncos",away="Green Bay Packers")

binary_cre

Internal function for binary model.

Description

An internal function.

Usage

binary_cre(Z_mat = Z_mat, first.order = first.order,
            home.field, control = control)

Arguments

Z_mat  data frame.
first.order  logical
home.field  logical
control  list
Description

2008 FBS College Football Regular Season Data

Usage

data(f2008)

Format

A data frame with 772 observations on the following 9 variables.

home  a factor
Game.Date  a POSIXlt date variable
away  a factor
home.response  a numeric vector
home.score  a numeric vector
away.response  a numeric vector
away.score  a numeric vector
neutral.site  a numeric vector
partition  a numeric vector

Source


Examples

data(f2008)
## maybe str(f2008) ; plot(f2008) ...
2009 FBS College Football Regular Season Data

Description

2009 FBS College Football Regular Season Data

Usage

data(f2009)

Format

A data frame with 772 observations on the following 7 variables.

- **home**: a factor
- **Game.Date**: a POSIXlt date variable
- **away**: a factor
- **home.response**: a numeric vector
- **home.score**: a numeric vector
- **away.response**: a numeric vector
- **away.score**: a numeric vector
- **neutral.site**: a numeric vector
- **partition**: a numeric vector

Source


Examples

data(f2009)
## maybe str(f2009) ; plot(f2009) ...
Description

2010 FBS College Football Regular Season Data

Usage

data(f2010)

Format

A data frame with 770 observations on the following 9 variables.

    home  a factor
    Game.Date  a POSIXlt
    away  a factor
    home.response  a numeric vector
    home.score  a numeric vector
    away.response  a numeric vector
    away.score  a numeric vector
    neutral.site  a numeric vector
    partition  a numeric vector

Source


Examples

data(f2010)
## maybe str(f2010); plot(f2010) ...
2011 FBS College Football Regular Season Data

Description

2011 FBS College Football Regular Season Data

Usage

data(f2011)

Format

A data frame with 781 observations on the following 9 variables.

- `home`: a factor
- `Game.Date`: a POSIXlt
- `away`: a factor
- `home.response`: a numeric vector
- `home.score`: a numeric vector
- `away.response`: a numeric vector
- `away.score`: a numeric vector
- `neutral.site`: a numeric vector
- `partition`: a numeric vector

Source


Examples

data(f2011)

## maybe str(f2011); plot(f2011) ...
Description

2012 FBS College Football Regular Season Data

Usage

data(f2012)

Format

A data frame with 809 observations on the following 9 variables.

home  a factor
Game.Date  a POSIXlt
away  a factor
home.response  a numeric vector
home.score  a numeric vector
away.response  a numeric vector
away.score  a numeric vector
neutral.site  a numeric vector
partition  a numeric vector

Source


Examples

data(f2012)
## maybe str(f2012) ; plot(f2012) ...
game.pred

Predict outcomes of games.

Description

After fitting a model with `mvglmmRank`, `game.pred` uses that model to predict outcomes of future matchups.

Usage

```r
game.pred(res, home, away, neutral.site = FALSE)
```

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>res</code></td>
<td>an object of class <code>mvglmmRank</code></td>
</tr>
<tr>
<td><code>home</code></td>
<td>a character string for the home team (use quotation marks!)</td>
</tr>
<tr>
<td><code>away</code></td>
<td>a character string for the away team (use quotation marks!)</td>
</tr>
<tr>
<td><code>neutral.site</code></td>
<td>logical. If TRUE, uses the neutral site mean score, assuming some of the games in the training data occurred at neutral sites.</td>
</tr>
</tbody>
</table>

Value

Prints predicted scores and/or predicted probability of a home team win, depending on the type of model specified by `res`.

Author(s)

Andrew T. Karl and Jennifer Broatch

References


Examples

data(nfl2012)
mvglmmRank(nfl2012,method="PB0",first.order=TRUE,verbose=TRUE,max.iter.EM=1)

result <- mvglmmRank(nfl2012,method="PB0",first.order=TRUE,verbose=TRUE)
print(result)
game.pred(result,home="Denver Broncos",away="Green Bay Packers")

Description

This function fits a (multivariate) generalized linear mixed model to team scores and/or win/loss indicators.

Usage

mvglmmRank(game.data, method = "PB0", first.order = FALSE, home.field = TRUE, max.iter.EM = 1000, tol1 = 1e-04, tol2 = 1e-04, tolFE = 0, tol.n = 1e-07, verbose = TRUE, OT.flag = FALSE, Hessian = FALSE, REML.N=TRUE)

Arguments

game.data a data frame that contains a column "home" of team names, a column "away" of team names, a column "home.response" containing the scores (or other response) of the "home" teams, a column "away.response" containing the scores (or other response) of the "away" teams, (optionally) a column "binary.response" that contains a column of binary responses (0's and 1's), and (optionally) a column "neutral.site" which takes the value 1 for neutral site games and 0 otherwise. NOTE: If game.data does not contain a "binary.response" column, then an indicator will be created for whether the home team won. NOTE: For neutral site games, randomly assign the teams as "home" or "away". As noted below, the data frame may optionally contain a column, OT, which indicates how many overtime periods were played. NOTE: the game.data$OT column should not contain missing data. If there was no overtime, specify "none" or 0.

method a character (remember to use quotation marks!). Choices are "N", "P0", "P1", "B", "NB", "PB0", "PB1", "NB.mov", or "N.mov". "N" indicates the scores are fit with a normal distribution with intra-game correlation between the home and away teams accounted for in an unstructured 2x2 error covariance matrix. "P" indicates the scores are fit with a Poisson distribution. "B" indicates the home win/loss indicators are fit using a binary distribution with a probit link. The presence of a "1" with a "P" indicates potential intra-game correlation is modeled with an additional game-level random effect. A "0" indicates no such random effects are included. "NB.mov" fits the margin of victory of the "home"
**Details**

Setting `first.order=TRUE` will yield the first order Laplace approximation. A partial fully exponential Laplace approximation can be obtained by setting `tol1 > tol2` and `tolFE=0`. This will apply fully exponential corrections to the vector of team ratings (the EBLUPs), but not to the covariance matrix of this vector. Karl, Yang, and Lohr (2014) show that this approach produces a large portion of the benefit of the fully exponential Laplace approximation in only a fraction of the time. Using the default tolerances of `mvglmmRank` leads to this behavior.

To summarize, the models (except for `method="N"`) run with the first order Laplace approximation until the relative change between parameters is <= `tol1`. If `first.order=TRUE`, the program stops.
Otherwise, the program continues with the Laplace approximation, applying fully exponential corrections to the random effects vector until the maximum of the relative parameter changes is <= tolFE. At this point, the program continues using the complete fully exponential Laplace approximation (corrections to both the random effects vector and its covariance matrix) until the maximum relative parameter change is <= tol2. If tolFE < tol2, then the program will finish without applying fully exponential corrections to the random effects covariance matrix.

method="PB1" is the least scalable, as the memory and computational requirements for this model are at least O((teams+number of games)^2). In the example data included with the package, the NCAA basketball data is slow with the fully exponential approximation and method="PB1".

Value

mvglmmRank returns an object of class mvglmmRank

An object of class mvglmmRank is a list containing the following components:

nratings.offense
The vector of offensive ratings from the normal model, or NULL if the normal model was not fit.
nratings.defense
The vector of defensive ratings from the normal model, or NULL if the normal model was not fit.
pratings.offense
The vector of offensive ratings from the Poisson model, or NULL if the Poisson model was not fit.
pratings.defense
The vector of defensive ratings from the Poisson model, or NULL if the Poisson model was not fit.
b.offense
The vector of win-propensity ratings from the binary model, or NULL if the binary model was not fit.
n.mean
Mean scores from the normal model.
p.mean
Mean scores from the Poisson model.
b.mean
Home field effect from the binary model.
G
Single block of random effects covariance matrix.
G.cor
Correlation matrix corresponding to covariance matrix G.
R
Error covariance matrix for normal model, or NULL if normal model not used.
R.cor
Error correlation matrix for normal model, or NULL if normal model not used.
home.field
Logical indicating whether or not a home field effect was modeled.
Hessian
The Hessian of the model parameters, if requested.
parameters
A vector of fitted model parameters.
N.output
NULL, or a list if method="N" or method="N.mov". In the later cases, the list contains the random effect design matrix Z, the fixed effects design matrix X, the estimated random effects covariance matrix G, the estimated error covariance matrix R, the predicted random effects eta, the joint covariance matrix of fixed and random effects ybetas_eblup_asycov, the covariance matrix of the fixed effects only ybetas_asycov, and the standard errors of the fixed effects ybetas_stderro.
fixed.effect.model.output

NULL, or a list if method="N.mov". In the later case, the list contains information about the results of fitting the margin of victory model with fixed (instead of random) team effects: the fixed effect design matrix $X$, the fixed effect parameter estimates $\beta$, logical indicating whether or not the home field effect is estimable is.mean.estimable (see Notes), the predicted margins of victory $\text{pred}$, the residuals $\text{resid}$, the fitted model variance $\sigma.sq$, and the covariance matrix of the random effects $\beta.covariance$. This can provide an unbiased estimate when the estimator from the mixed model is biased (Karl and Zimmerman, 2021).

The function game.pred may be used to predict the outcome of future games.

Author(s)

Andrew T. Karl <akarl@asu.edu>, Jennifer Broatch

References


See Also

See also game.pred

Examples

data(nfl2012)
mvglmmRank(nfl2012,method="PB0",first.order=TRUE,verbose=TRUE,max.iter.EM=1)

result <- mvglmmRank(nfl2012,method="PB0",first.order=TRUE,verbose=TRUE)
print(result)
game.pred(result,home="Denver Broncos",away="Green Bay Packers")
nba2013

2013 NBA Data

Description

2013 NBA Data

Usage

data(nba2013)

Format

A data frame with 1229 observations on the following 11 variables.

Date  a factor
away  a factor
home  a factor
OT  a factor
partition  a numeric vector
neutral.site  a numeric vector
ot.count  a numeric vector
home.response  a numeric vector
home.score  a numeric vector
away.response  a numeric vector
away.score  a numeric vector

Source

http://masseyratings.com/data.php

Examples

data(nba2013)
## maybe str(nba2013); plot(nba2013) ...
NB_crep

Internal Function for Normal-Binary Model

Description
Internal Function for Normal-Binary Model

Usage
NB_crep(Z_mat = Z_mat, first.order = first.order,
         home.field = home.field, control = control)

Arguments
Z_mat     data frame
first.order   logical
home.field   logical
control     list

NB_mov

Internal Function for Normal-Binary Model

Description
Internal Function for Normal-Binary Model

Usage
NB_mov(Z_mat = Z_mat, first.order = first.order,
       home.field = home.field, control = control)

Arguments
Z_mat     data frame
first.order   logical
home.field   logical
control     list
ncaab2012 2012 NCAA Division I Basketball Results

Description

2012 NCAA Division I Basketball Results

Usage

data(ncaab2012)

Format

A data frame with 5253 observations on the following 10 variables.

date  a factor
away  a factor
home  a factor
neutral.site  a numeric vector
partition  a numeric vector
home_win  a numeric vector
home.response  a numeric vector
home.score  a numeric vector
away.response  a numeric vector
away.score  a numeric vector

Source

http://masseyratings.com/data.php

Examples

data(ncaab2012)
## maybe str(ncaab2012) ; plot(ncaab2012) ...
Description

2012 NFL Regular Season Data

Usage

data(nfl2012)

Format

A data frame with 256 observations on the following 9 variables.

Date  a factor
away  a factor
home  a factor
neutral.site  a numeric vector
home.response  a numeric vector
home.score  a numeric vector
away.response  a numeric vector
away.score  a numeric vector
partition  a numeric vector

Source

http://masseyratings.com/data.php

Examples

data(nfl2012)
## maybe str(nfl2012) ; plot(nfl2012) ...
normal_cre

**Internal Function for Normal Model**

**Description**
Internal Function for Normal Model

**Usage**

```r
normal_cre(Z_mat = Z_mat, first.order = first.order,
            home.field = home.field, control = control)
```

**Arguments**

- `Z_mat` : data frame
- `first.order` : logical
- `home.field` : logical
- `control` : list

---

N_mov

**Internal Function for Normal MOV model**

**Description**
Internal Function for Normal MOV model

**Usage**

```r
N_mov(Z_mat = Z_mat, first.order = TRUE,
      home.field = home.field, control = control)
```

**Arguments**

- `Z_mat` : data frame
- `first.order` : logical
- `home.field` : logical
- `control` : list
PB_cre

Internal Function for Poisson-binary Model

Description

Internal Function for Poisson-binary Model

Usage

PB_cre(Z_mat = Z_mat, first.order = first.order,
       home.field = home.field, control = control,
       game.effect = game.effect)

Arguments

- **Z_mat**: data frame
- **first.order**: logical
- **home.field**: logical
- **control**: list
- **game.effect**: logical

poisson_cre

Internal Function for Poisson Model

Description

Internal Function for Poisson Model

Usage

poisson_cre(Z_mat = Z_mat, first.order = first.order,
            control = control, game.effect = game.effect,
            home.field = home.field)

Arguments

- **Z_mat**: data frame
- **first.order**: logical
- **control**: logical
- **game.effect**: logical
- **home.field**: logical
Index

* datasets
  - f2008, 4
  - f2009, 5
  - f2010, 6
  - f2011, 7
  - f2012, 8
  - nba2013, 14
  - ncaa2012, 16
  - nfl2012, 17

* package
  - mvglmmRank-package, 2

* regression
  - game.pred, 9
  - mvglmmRank, 10

binary_cre, 3

f2008, 4
f2009, 5
f2010, 6
f2011, 7
f2012, 8

game.pred, 9, 13

mvglmmRank, 9, 10
mvglmmRank-package, 2

N_mov, 18
NB_cre, 15
NB_mov, 15
nba2013, 14
ncaa2012, 16
nfl2012, 17
normal_cre, 18

PB_cre, 19
poisson_cre, 19