Package 'footBayes'

October 7, 2026

Type Package

Title Fitting Bayesian and MLE Football Models

Version 2.1.0

Date 2026-09-07

Maintainer Leonardo Egidi <legidi@units.it>

License GPL-2

Description This is the first package allowing for the estimation, visualization and prediction of the most well-known football models: double Poisson, bivariate Poisson, Dixon-Coles, negative binomial, Skellam, student_t, diagonal-inflated bivariate Poisson, and zero-inflated Skellam. It supports both maximum likelihood estimation (MLE, for 'static' models only) and Bayesian inference. Team abilities can be static or dynamic over weeks or seasons, with alternative specifications of the evolution variance: a common variance (Owen, 2011), variance inflation after the summer break (Koopman and Lit, 2015) <doi:10.1111/rssa.12042>, and weighted dynamic models with commensurate priors (Macrì-Demartino, Egidi and Torelli, 2026) <doi:10.1093/jrsssc/qlag032>. Historical team strengths can be estimated through a Bayesian Bradley-Terry-Davidson model and used as a covariate. For Bayesian methods, it incorporates several techniques: MCMC sampling with Hamiltonian Monte Carlo, variational inference using either the Pathfinder algorithm or Automatic Differentiation Variational Inference (ADVI), and the Laplace approximation. The package compiles all the 'CmdStan' models once during installation using the 'instantiate' package. The model construction relies on the most well-known football references, such as Dixon and Coles (1997) <doi:10.1111/1467-9876.00065>, Karlis and Ntzoufras (2003) <doi:10.1111/1467-9884.00366> and Egidi, Pauli and Torelli (2018) <doi:10.1177/1471082X18798414>.

URL https://github.com/leoegidi/footbayes

BugReports https://github.com/leoegidi/footbayes/issues

Encoding UTF-8

SystemRequirements CmdStan (https://mc-stan.org/users/interfaces/cmdstan), pandoc (>= 1.12.3), pandoc-citeproc

Depends R (>= 4.2.0)

Imports rstan (>= 2.18.1), instantiate, reshape2, ggplot2, ggridges, matrixStats, extraDistr, metRology, dplyr, tidyr, numDeriv, magrittr, rlang, posterior

Suggests testthat (>= 3.0.0), knitr (>= 1.37), rmarkdown (>= 2.10), loo, bayesplot, cmdstanr (>= 0.6.0)

Additional_repositories https://stan-dev.r-universe.dev/

VignetteBuilder knitr

LazyData true

BuildManual yes

Config/testthat/edition 3

Config/roxygen2/version 8.1.0

NeedsCompilation yes

Author Leonardo Egidi [aut, cre], Roberto Macrì Demartino [aut], Vasilis Palaskas [aut]

Repository CRAN

Date/Publication 2026-10-07 12:50:28 UTC

Contents

btd_foot . . . 3
compare_foot . . . 6
england . . . 8
foot_abilities . . . 9
foot_prob . . . 11
foot_rank . . . 13
foot_round_robin . . . 14
italy . . . 16
mle_foot . . . 17
plot_btdPosterior . . . 20
plot_logStrength . . . 22
pp_foot . . . 23
print.btdFoot . . . 25
print.compareFoot . . . 26
print.stanFoot . . . 27
priors . . . 27
stan_foot . . . 29

Index **37**

btd_foot *Bayesian Bradley-Terry-Davidson Model*

Description

Fits a Bayesian Bradley-Terry-Davidson model using Stan. Supports both static and dynamic ranking models, allowing for the estimation of team strengths over time.

Usage

```
btd_foot(
  data,
  dynamic_rank = FALSE,
  home_effect = FALSE,
 prior_par = list(logStrength = normal(0, 3), logTie = normal(0, 0.3), home = normal(0,
    5)),
  rank_measure = "median",
  method = "MCMC",
  ...
)
```

Arguments

data	A data frame containing the observations with columns: • periods: Time point of each observation (integer >= 1). • home_team: Home team's name (character string). • away_team: Away team's name (character string). • match_outcome: Outcome (1 if home team beats away team, 2 for tie, and 3 if away team beats home team). The data frame must not contain missing values.
dynamic_rank	A logical value indicating whether a dynamic ranking model is used (default is FALSE).
home_effect	A logical value indicating the inclusion of a home effect in the model. (default is FALSE).
prior_par	A list specifying the prior distributions for the parameters of interest, using the normal function: • logStrength: Prior for the team log-strengths. Default is normal(0, 3). • logTie: Prior for the tie parameter. Default is normal(0, 0.3). • home: Prior for the home effect (home). Applicable only if home_effect = TRUE. Default is normal(0, 5). Only normal priors are allowed for this model.
rank_measure	A character string specifying the method used to summarize the posterior distributions of the team strengths. Options are:

- "median": Uses the median of the posterior samples (default).
- "mean": Uses the mean of the posterior samples.
- "map": Uses the Maximum A Posteriori estimate, calculated as the mode of the posterior distribution.

method — A character string specifying the method used to obtain the Bayesian estimates. Options are:

- "MCMC": Markov chain Monte Carlo algorithm (default).
- "VI": Automatic differentiation variational inference algorithms.
- "pathfinder": Pathfinder variational inference algorithm.
- "laplace": Laplace algorithm.

... — Additional arguments passed to cmdstanr (e.g., iter_sampling, chains, parallel_chains).

Value

An object of class "btdFoot", which is a list containing:

- fit: The fitted CmdStanFit object returned by cmdstanr.
- rank: A data frame with the rankings, including columns:
  - periods: The time period.
  - team: The team name.
  - log_strengths: The estimated log-strength of the team based on the chosen rank_measure.
- data: The input data.
- stan_data: The data list passed to Stan.
- stan_code: The Stan code of the underline model.
- stan_args: The optional cmdstanr parameters passed to (...).
- rank_measure: The summary statistic used to compute the rankings.
- alg_method: The inference algorithm used to obtain the Bayesian estimates.

Author(s)

Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>.

Examples

```
## Not run:
if (instantiate::stan_cmdstan_exists()) {
  library(dplyr)

  data("italy")

  italy_2020_2021 <- italy %>%
    dplyr::select(Season, home, visitor, hgoal, vgoal) %>%
    dplyr::filter(Season == "2020" | Season == "2021") %>%
    dplyr::mutate(match_outcome = dplyr::case_when(
      hgoal > vgoal ~ 1, # Home team wins
      hgoal == vgoal ~ 2, # Draw
```

```
      hgoal < vgoal ~ 3 # Away team wins
    )) %>%
    dplyr::mutate(periods = dplyr::case_when(
      dplyr::row_number() <= 190 ~ 1,
      dplyr::row_number() <= 380 ~ 2,
      dplyr::row_number() <= 570 ~ 3,
      TRUE ~ 4
    )) %>% # Assign periods based on match number
    dplyr::select(periods,
      home_team = home,
      away_team = visitor, match_outcome
    )

  # Dynamic Ranking Example with Median Rank Measure
  fit_result_dyn <- btd_foot(
    data = italy_2020_2021,
    dynamic_rank = TRUE,
    home_effect = TRUE,
    prior_par = list(
      logStrength = normal(0, 10),
      logTie = normal(0, 5),
      home = normal(0, 5)
    ),
    rank_measure = "median",
    iter_sampling = 1000,
    parallel_chains = 2,
    chains = 2
  )

  print(fit_result_dyn)

  print(fit_result_dyn, pars = c("logStrength", "home"), teams = c("AC Milan", "AS Roma"))

  # Static Ranking Example with MAP Rank Measure
  fit_result_stat <- btd_foot(
    data = italy_2020_2021,
    dynamic_rank = FALSE,
    prior_par = list(
      logStrength = normal(0, 10),
      logTie = normal(0, 5),
      home = normal(0, 5)
    ),
    rank_measure = "map",
    iter_sampling = 1000,
    parallel_chains = 2,
    chains = 2
  )

  print(fit_result_stat)
}

## End(Not run)
```

compare_foot *Compare Football Models using Various Metrics*

Description

Compares multiple football models or directly provided probability matrices based on specified metrics (accuracy, Brier score, ranked probability score, Pseudo R^2, average of correct probabilities), using a test dataset. Additionally, computes the confusion matrices. The function returns an object of class `compareFoot`.

Usage

```
compare_foot(
  source,
  test_data,
  metric = c("accuracy", "brier", "ACP", "pseudoR2", "RPS"),
  conf_matrix = FALSE
)
```

Arguments

`source`	A named list containing either: • Fitted model objects (of class `stanFoot`, `CmdStanFit`, `stanfit`), each representing a football model. • Matrices where each matrix contains the estimated probabilities for "Home Win," "Draw," and "Away Win" in its columns.
`test_data`	A data frame containing the test dataset, with columns: • `home_team`: Home team's name (character string). • `away_team`: Away team's name (character string). • `home_goals`: Goals scored by the home team (integer >= 0). • `away_goals`: Goals scored by the away team (integer >= 0).
`metric`	A character vector specifying the metrics to use for comparison. Options are: • `"accuracy"`: Computes the accuracy of each model. • `"brier"`: Computes the Brier score of each model. • `"RPS"`: Computes the ranked probability score (RPS) for each model. • `"ACP"`: Computes the average of correct probabilities (ACP) for each model. • `"pseudoR2"`: Computes the Pseudo R^2, defined as the geometric mean of the probabilities assigned to the actual results. Default is `c("accuracy", "brier", "ACP", "pseudoR2", "RPS")`, computing the specified metrics.
`conf_matrix`	A logical value indicating whether to generate a confusion matrix comparing predicted outcomes against actual outcomes for each model or probability matrix. Default is `FALSE`.

Details

The function extracts predictions from each model or directly uses the provided probability matrices and computes the chosen metrics on the test dataset. It also possible to compute confusion matrices.

Value

An object of class `compareFoot`, which is a list containing:

- `metrics`: A data frame containing the metric values for each model or probability matrix.
- `confusion_matrix`: Confusion matrices for each model or probability matrix.

Author(s)

Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>

Examples

```
## Not run:
if (instantiate::stan_cmdstan_exists()) {
  library(dplyr)

  data("italy")
  italy_2000 <- italy %>%
    dplyr::select(Season, home, visitor, hgoal, vgoal) %>%
    dplyr::filter(Season == "2000")

 colnames(italy_2000) <- c("periods", "home_team", "away_team", "home_goals", "away_goals")

  # Example with fitted models
  fit_1 <- stan_foot(
    data = italy_2000,
    model = "double_pois", predict = 18
  ) # Double Poisson model
  fit_2 <- stan_foot(
    data = italy_2000,
    model = "biv_pois", predict = 18
  ) # Bivariate Poisson model

  italy_2000_test <- italy_2000[289:306, ]


  compare_results_models <- compare_foot(
    source = list(
      double_poisson = fit_1,
      bivariate_poisson = fit_2
    ),
    test_data = italy_2000_test,
    metric = c("accuracy", "brier", "ACP", "pseudoR2", "RPS"),
    conf_matrix = TRUE
  )

  print(compare_results_models)
```

```
  # Example with probability matrices

  home_team <- c(
    "AC Milan", "Inter", "Juventus", "AS Roma", "Napoli",
    "Lazio", "Atalanta", "Fiorentina", "Torino", "Sassuolo", "Udinese"
  )

  away_team <- c(
    "Juventus", "Napoli", "Inter", "Atalanta", "Lazio",
    "AC Milan", "Sassuolo", "Torino", "Fiorentina", "Udinese", "AS Roma"
  )

  # Home and Away goals based on given data
  home_goals <- c(2, 0, 2, 2, 3, 1, 4, 2, 1, 1, 2)
  away_goals <- c(1, 0, 1, 3, 2, 1, 1, 2, 1, 1, 2)

  # Combine into a data frame
  test_data <- data.frame(home_team, away_team, home_goals, away_goals)

  # Define the data for each column
  pW <- c(0.51, 0.45, 0.48, 0.53, 0.56, 0.39, 0.52, 0.55, 0.61, 0.37, 0.35)
  pD <- c(0.27, 0.25, 0.31, 0.18, 0.23, 0.30, 0.24, 0.26, 0.18, 0.19, 0.22)
  pL <- c(0.22, 0.30, 0.21, 0.29, 0.21, 0.31, 0.24, 0.19, 0.21, 0.44, 0.43)

  # Create the data frame table_prob
  table_prob <- data.frame(pW, pD, pL)
  matrix_prob <- as.matrix(table_prob)

  # Use compare_foot function
  compare_results_matrices <- compare_foot(
    source = list(matrix_1 = matrix_prob),
    test_data = test_data,
    metric = c("accuracy", "brier", "pseudoR2", "ACP", "RPS")
  )
  # Print the results
  print(compare_results_matrices)
}

## End(Not run)
```

`england` *English league results 1888-2022*

Description

All results for English soccer games in the top 4 tiers from 1888/89 season to 2021/22 season.

Usage

```
england
```

Format

A data frame with 203956 rows and 12 variables:

Date Date of match

Season Season of match - refers to starting year

home Home team

visitor Visiting team

FT Full-time result

hgoal Goals scored by home team

vgoal Goals scored by visiting team

division Division: 1,2,3,4 or 3N (Old 3-North) or 3S (Old 3-South)

tier Tier of football pyramid: 1,2,3,4

totgoal Total goals in game

goaldif Goal difference in game home goals - visitor goals

result Result: H-Home Win, A-Away Win, D-Draw

---

foot_abilities *Plot football abilities from Stan and MLE models*

Description

Depicts teams' abilities either from the Stan models fitted via the `stan_foot` function or from MLE models fitted via the `mle_foot` function. Supported MLE models include `"double_pois"`, `"biv_pois"`, `"dixon_coles"`, `"neg_bin"`, `"skellam"`, and `"student_t"`. Models with separate attack and defence parameters (all except `"student_t"`) display attack and defence effects; the `"student_t"` model displays combined global abilities.

Usage

```
foot_abilities(object, data, type = "both", teams = NULL)
```

Arguments

`object`	An object either of class `stanFoot`, `CmdStanFit`, `stanfit`, or class `list` containing the Maximum Likelihood Estimates (MLE) for the model parameters fitted with `mle_foot`.
`data`	A data frame containing match data with columns: • `periods`: Time point of each observation (integer >= 1). • `home_team`: Home team's name (character string).

- away_team: Away team's name (character string).
- home_goals: Goals scored by the home team (integer >= 0).
- away_goals: Goals scored by the away team (integer >= 0).

type	Type of ability for models with separate attack/defence parameters: one among "defense", "attack" or "both". Default is "both". Ignored for the "student_t" model which uses combined abilities.
teams	An optional character vector specifying team names to include. If NULL, all teams are included.

Value

A ggplot object showing each selected team's ability estimates:

- For static Bayesian or MLE fits, horizontal error bars (95% intervals) and point estimates.
- For dynamic Bayesian fits, ribbon and line plots over periods.

Author(s)

Leonardo Egidi <legidi@units.it> and Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>.

Examples

```
## Not run:
if (instantiate::stan_cmdstan_exists()) {
  library(dplyr)

  data("italy")
  italy <- as_tibble(italy)

  ### no dynamics, no prediction

  italy_2000_2002 <- italy %>%
    dplyr::select(Season, home, visitor, hgoal, vgoal) %>%
    dplyr::filter(Season == "2000" | Season == "2001" | Season == "2002")

 colnames(italy_2000_2002) <- c("periods", "home_team", "away_team", "home_goals", "away_goals")

  fit1 <- stan_foot(
    data = italy_2000_2002,
    model = "double_pois"
  ) # double poisson

  fit2 <- stan_foot(
    data = italy_2000_2002,
    model = "biv_pois"
  ) # bivariate poisson

  fit3 <- stan_foot(
    data = italy_2000_2002,
    model = "skellam"
  ) # skellam
```

```
  fit4 <- stan_foot(
    data = italy_2000_2002,
    model = "student_t"
  ) # student_t

  foot_abilities(fit1, italy_2000_2002)
  foot_abilities(fit2, italy_2000_2002)
  foot_abilities(fit3, italy_2000_2002)
  foot_abilities(fit4, italy_2000_2002)

  ### seasonal dynamics, predict the last season

  fit5 <- stan_foot(
    data = italy_2000_2002,
    model = "biv_pois",
    predict = 180,
    dynamic_type = "seasonal"
  ) # bivariate poisson
  foot_abilities(fit5, italy_2000_2002)
}

### MLE models (no Stan required)

data("italy")
italy_2000 <- italy[italy$Season == "2000", ]
italy_2000 <- data.frame(
  periods = italy_2000$Season,
  home_team = italy_2000$home,
  away_team = italy_2000$visitor,
  home_goals = italy_2000$hgoal,
  away_goals = italy_2000$vgoal
)

## Dixon-Coles MLE
mle_dc <- mle_foot(data = italy_2000, model = "dixon_coles")
foot_abilities(mle_dc, italy_2000)

## Negative Binomial MLE
mle_nb <- mle_foot(data = italy_2000, model = "neg_bin")
foot_abilities(mle_nb, italy_2000, type = "attack")

## End(Not run)
```

`foot_prob` *Plot football matches probabilities for out-of-sample football matches.*

Description

The function provides a table containing the home win, draw and away win probabilities for a bunch of out-of-sample matches as specified by `stan_foot` or `mle_foot`.

Usage

```
foot_prob(object, data, home_team, away_team)
```

Arguments

object	An object either of class stanFoot, CmdStanFit, stanfit, or class list containing the Maximum Likelihood Estimates (MLE) for the model parameters fitted with mle_foot.
data	A data frame containing match data with columns: • periods: Time point of each observation (integer >= 1). • home_team: Home team's name (character string). • away_team: Away team's name (character string). • home_goals: Goals scored by the home team (integer >= 0). • away_goals: Goals scored by the away team (integer >= 0).
home_team	The home team(s) for the predicted matches.
away_team	The away team(s) for the predicted matches.

Details

For Bayesian models the results probabilities are computed according to the simulation from the posterior predictive distribution of future (out-of-sample) matches. Specifically, matches are ordered from those in which the favorite team has the highest posterior probability of winning to those where the underdog is more likely to win. For MLE models fitted via the mle_foot the probabilities are computed by simulating from the MLE estimates. Supported MLE models include: "double_pois", "biv_pois", "dixon_coles", "neg_bin", "skellam", and "student_t".

For the Dixon-Coles model, predictions incorporate the low-score dependence adjustment via weighted resampling from independent Poisson simulations. For the negative binomial model, predictions use the NB2 parameterization with separate overdispersion parameters for home and away goals.

Value

A list with components:

- prob_table: A data frame containing the results probabilities of the out-of-sample matches.
- prob_plot: A ggplot object for Bayesian models only showing the posterior predictive heatmap of exact score probabilities, with the true result highlighted.

Author(s)

Leonardo Egidi <legidi@units.it> and Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>.

Examples

```
## Not run:
if (instantiate::stan_cmdstan_exists()) {
  library(dplyr)
```

```
data("italy")
italy_2000 <- italy %>%
  dplyr::select(Season, home, visitor, hgoal, vgoal) %>%
  dplyr::filter(Season == "2000")

colnames(italy_2000) <- c("periods", "home_team", "away_team", "home_goals", "away_goals")



fit <- stan_foot(
  data = italy_2000,
  model = "double_pois",
  predict = 18
) # double pois

foot_prob(
  fit, italy_2000, "Inter",
  "Bologna FC"
)

foot_prob(fit, italy_2000) # all the out-of-sample matches
}

## End(Not run)
```

`foot_rank` *Rank and points predictions*

Description

Posterior predictive plots and final rank table for football seasons.

Usage

```
foot_rank(object, data, teams = NULL, visualize = "individual")
```

Arguments

`object`	An object either of class `stanFoot`, `CmdStanFit`, or `stanfit`.
`data`	A data frame containing match data with columns: • `periods`: Time point of each observation (integer >= 1). • `home_team`: Home team's name (character string). • `away_team`: Away team's name (character string). • `home_goals`: Goals scored by the home team (integer >= 0). • `away_goals`: Goals scored by the away team (integer >= 0).
`teams`	An optional character vector specifying team names to include. If `NULL`, all teams are included.
`visualize`	Type of plots, one among `"aggregated"` or `"individual"`. Default is `"individual"`.

Details

For Bayesian models fitted via `stan_foot` the final rank tables are computed according to the simulation from the posterior predictive distribution of future (out-of-sample) matches. The dataset should refer to one or more seasons from a given national football league (Premier League, Serie A, La Liga, etc.).

Value

If `visualize = "aggregated"`: a list with

- `rank_table`: A data frame of observed and simulated final points (median, 25%/75% quantiles).
- `rank_plot`: A `ggplot` comparing observed vs simulated final points for each team.

If `visualize = "individual"`: A `ggplot` showing, for each selected team, the observed and simulated cumulative points over match-days.

Author(s)

Leonardo Egidi <legidi@units.it> and Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>

Examples

```
## Not run:
if (instantiate::stan_cmdstan_exists()) {
  library(dplyr)

  data("italy")
  italy_1999_2000 <- italy %>%
    dplyr::select(Season, home, visitor, hgoal, vgoal) %>%
    dplyr::filter(Season == "1999" | Season == "2000")

 colnames(italy_1999_2000) <- c("periods", "home_team", "away_team", "home_goals", "away_goals")

  fit <- stan_foot(italy_1999_2000, "double_pois", iter_sampling = 200)
  foot_rank(fit, italy_1999_2000)
  foot_rank(fit, italy_1999_2000, visualize = "individual")
}

## End(Not run)
```

`foot_round_robin` *Round-robin for football leagues*

Description

Posterior predictive probabilities for a football season in a round-robin format

Usage

```
foot_round_robin(object, data, teams = NULL, output = "both")
```

Arguments

object	An object either of class stanFoot, CmdStanFit, stanfit.
data	A data frame containing match data with columns: • periods: Time point of each observation (integer >= 1). • home_team: Home team's name (character string). • away_team: Away team's name (character string). • home_goals: Goals scored by the home team (integer >= 0). • away_goals: Goals scored by the away team (integer >= 0).
teams	An optional character vector specifying team names to include. If NULL, all teams are included.
output	An optional character string specifying the type of output to return. One of "both", "table", or "plot". Default is "both".

Details

For Bayesian models fitted via stan_foot the round-robin table is computed according to the simulation from the posterior predictive distribution of future (out-of-sample) matches. The dataset should refer to one or more seasons from a given national football league (Premier League, Serie A, La Liga, etc.).

Value

If output = "both" a list with:

- round_table: A data frame of matchups (Home, Away), observed scores, and Home_prob (median posterior probability of a home win).
- round_plot: A ggplot heatmap of home-win probabilities with observed scores overlaid.

If output = "table" or "plot", returns only that component.

Author(s)

Leonardo Egidi <legidi@units.it> and Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>

Examples

```
## Not run:
if (instantiate::stan_cmdstan_exists()) {
  library(dplyr)

  data("italy")
  italy_1999_2000 <- italy %>%
    dplyr::select(Season, home, visitor, hgoal, vgoal) %>%
    dplyr::filter(Season == "1999" | Season == "2000")
```

```
 colnames(italy_1999_2000) <- c("periods", "home_team", "away_team", "home_goals", "away_goals")

  fit <- stan_foot(italy_1999_2000, "double_pois", predict = 45, iter_sampling = 200)

  foot_round_robin(fit, italy_1999_2000)
  foot_round_robin(fit, italy_1999_2000, c("Parma AC", "AS Roma"))
}

## End(Not run)
```

`italy` *Italy league results 1929-2022*

Description

All results for Italian soccer games in the top tier from 1929/30 season to 2021/22 season.

Usage

```
italy
```

Format

A data frame with 27684 rows and 8 variables:

Date Date of match

Season Season of match - refers to starting year

home Home team

visitor Visiting team

FT Full-time result

hgoal Goals scored by home team

vgoal Goals scored by visiting team

tier Tier of football pyramid: 1

`mle_foot` *Fit football models with Maximum Likelihood*

Description

Fits football goal-based models using maximum likelihood estimation. Supported models include: double Poisson, bivariate Poisson, Dixon-Coles, negative binomial, Skellam, and Student's t.

Usage

```
mle_foot(
  data,
  model,
  predict = 0,
  maxit = 1000,
  method = "BFGS",
  interval = "profile",
  hessian = FALSE,
  sigma_y = 1
)
```

Arguments

`data`	A data frame containing match data with columns: • `periods`: Time point of each observation (integer >= 1). • `home_team`: Home team's name (character string). • `away_team`: Away team's name (character string). • `home_goals`: Goals scored by the home team (integer >= 0). • `away_goals`: Goals scored by the away team (integer >= 0).
`model`	A character specifying the model to fit. Options are: • `"double_pois"`: Double Poisson model. • `"biv_pois"`: Bivariate Poisson model. • `"dixon_coles"`: Dixon-Coles model. • `"neg_bin"`: Negative Binomial model. • `"skellam"`: Skellam model. • `"student_t"`: Student's t model.
`predict`	An integer specifying the number of out-of-sample matches for prediction. If missing, the function fits the model to the entire dataset without making predictions.
`maxit`	An integer specifying the maximum number of optimizer iterations default is 1000).
`method`	A character specifying the optimization method. Options are • `"Nelder-Mead"`.

- "BFGS" (default).
- "CG".
- "L-BFGS-B".
- "SANN".
- "Brent".

For further details see {optim} function in stats package.

interval — A character specifying the interval type for confidence intervals. Options are

- "profile" (default).
- "Wald".

hessian — A logical value indicating to include the computation of the Hessian (default FALSE).

sigma_y — A positive numeric value indicating the scale parameter for Student t likelihood (default 1).

Details

MLE can be obtained only for static models, with no time-dependence. Likelihood optimization is performed via the BFGS method of the {optim} function in stats package.

Value

A named list containing:

- att: A matrix of attack ratings, with MLE and 95% confidence intervals (for "double_pois", "biv_pois", "dixon_coles", "neg_bin" and "skellam" models).
- def: A matrix of defence ratings, with MLE and 95% confidence intervals (for "double_pois", "biv_pois", "dixon_coles", "neg_bin" and "skellam" models).
- abilities: A matrix of combined ability, with MLE and 95% confidence intervals (for "student_t" only).
- home_effect: A matrix with with MLE and 95% confidence intervals for the home effect estimate.
- corr: A matrix with MLE and 95% confidence intervals for the bivariate Poisson correlation parameter (for "biv_pois" only).
- rho: A matrix with MLE and 95% confidence intervals for the Dixon-Coles dependence parameter (for "dixon_coles" only).
- overdispersion: A matrix with MLE and 95% confidence intervals for the home and away overdispersion parameters (for "neg_bin" only).
- model: The name of the fitted model (character).
- predict: The number of out-of-sample matches used for prediction (integer).
- sigma_y: The scale parameter used in the Student t likelihood (for "student_t" only).
- team1_prev: Integer indices of home teams in the out-of-sample matches (if predict > 0).
- team2_prev: Integer indices of away teams in the out-of-sample matches (if predict > 0).
- logLik: The maximized log likelihood (numeric).
- aic: Akaike Information Criterion (numeric).
- bic: Bayesian Information Criterion (numeric).

Author(s)

Leonardo Egidi <legidi@units.it> and Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>

References

Baio, G. and Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. Journal of Applied Statistics 37(2), 253-264.

Dixon, M. J. and Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. Journal of the Royal Statistical Society: Series C (Applied Statistics) 46(2), 265-280.

Egidi, L., Pauli, F., and Torelli, N. (2018). Combining historical data and bookmakers' odds in modelling football scores. Statistical Modelling, 18(5-6), 436-459.

Gelman, A. (2014). Stan goes to the World Cup. From "Statistical Modeling, Causal Inference, and Social Science" blog.

Karlis, D. and Ntzoufras, I. (2003). Analysis of sports data by using bivariate poisson models. Journal of the Royal Statistical Society: Series D (The Statistician) 52(3), 381-393.

Karlis, D. and Ntzoufras,I. (2009). Bayesian modelling of football outcomes: Using the Skellam's distribution for the goal difference. IMA Journal of Management Mathematics 20(2), 133-145.

Owen, A. (2011). Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. IMA Journal of Management Mathematics, 22(2), 99-113.

Examples

```
## Not run:
library(dplyr)

data("italy")
italy <- as_tibble(italy)
italy_2000_2002 <- italy %>%
  dplyr::select(Season, home, visitor, hgoal, vgoal) %>%
  dplyr::filter(Season == "2000" | Season == "2001" | Season == "2002")

colnames(italy_2000_2002) <- c("periods", "home_team", "away_team", "home_goals", "away_goals")

mle_fit <- mle_foot(
  data = italy_2000_2002,
  model = "double_pois"
)

## End(Not run)
```

plot_btdPosterior *Plot Posterior Distributions for* btdFoot *Objects*

Description

Plots for the posterior distributions of team log-strengths and other parameters with customizable plot types and facets.

Usage

```
plot_btdPosterior(
  x,
  pars = "logStrength",
  plot_type = "boxplot",
  teams = NULL,
  ncol = NULL,
  scales = NULL
)
```

Arguments

x	An object of class btdFoot.
pars	A character string specifying the parameter to plot. Choices are "logStrength", "logTie", and "home". Default is "logStrength".
plot_type	A character string specifying the type of plot. Choices are "boxplot" and "density". Default is "boxplot".
teams	An optional character vector specifying team names to include in the posterior boxplots or density plots. If NULL, all teams are included.
ncol	An optional integer specifying the number of columns in the facet wrap when using a dynamic Bayesian Bradley-Terry-Davidson model. Default is 8.
scales	An optional character string specifying the scales for the facets when using a dynamic Bayesian Bradley-Terry-Davidson model. Options include "free", "fixed", "free_x", and "free_y". Default is "free_x".

Value

A ggplot object displaying:

- For pars="logStrength":
  - Dynamic BTD: Faceted boxplots or density plots (including the 95% credible interval) of posterior log-strengths by team and period.
  - Static BTD: Boxplots or density plots (including the 95% credible interval) of posterior log-strengths for each team.
- For pars="logTie" or pars="home": A single boxplot or density plot with 95% credible interval.

Author(s)

Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>.

Examples

```
## Not run:
if (instantiate::stan_cmdstan_exists()) {
  library(dplyr)

  # Load example data
  data("italy")

  # Prepare the data
  italy_2020_2021_rank <- italy %>%
    select(Season, home, visitor, hgoal, vgoal) %>%
    filter(Season %in% c("2020", "2021")) %>%
    mutate(match_outcome = case_when(
      hgoal > vgoal ~ 1, # Home team wins
      hgoal == vgoal ~ 2, # Draw
      hgoal < vgoal ~ 3 # Away team wins
    )) %>%
    mutate(periods = case_when(
      row_number() <= 190 ~ 1,
      row_number() <= 380 ~ 2,
      row_number() <= 570 ~ 3,
      TRUE ~ 4
    )) %>% # Assign periods based on match number
    select(periods,
      home_team = home,
      away_team = visitor, match_outcome
    )

  # Fit the Bayesian Bradley-Terry-Davidson model with dynamic ranking
  fit_rank_dyn <- btd_foot(
    data = italy_2020_2021_rank,
    dynamic_rank = TRUE,
    rank_measure = "median",
    iter_sampling = 1000,
    parallel_chains = 2,
    chains = 2
  )

  # Plot posterior distributions with default settings
  plot_btdPosterior(fit_rank_dyn)

  # Plot posterior distributions for specific teams with customized facets
  plot_btdPosterior(
    fit_rank_dyn,
    teams = c("AC Milan", "AS Roma", "Juventus", "Inter"),
    ncol = 2
  )
```

```
  plot_btdPosterior(
    fit_rank_dyn,
    plot_type = "density",
    teams = c("AC Milan", "AS Roma", "Juventus", "Inter"),
    ncol = 2
  )
}

## End(Not run)
```

`plot_logStrength` *Plot Rankings for btdFoot Objects*

Description

Visualizes team rankings based on whether the ranking is dynamic or static.

Usage

```
plot_logStrength(x, teams = NULL)
```

Arguments

`x`	An object of class `btdFoot`.
`teams`	An optional character vector specifying team names to include in the rankings plot. If `NULL`, all teams are included.

Details

- Dynamic Ranking: Plots Rank Points over Periods for each team with lines and points.
- Static Ranking: Plots Rank Points on the x-axis against Team Names on the y-axis with horizontal lines and points.

Value

A `ggplot` object:

- Dynamic BTD: A lineplot for the `log_strengths` over each period, colored by team.
- Static BTD: A horizontal segment-and-dot plot of the `log_strengths`, one row per team, sorted by log-strength.

Author(s)

Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>.

Examples

```
## Not run:
if (instantiate::stan_cmdstan_exists()) {
  library(dplyr)

  data("italy")

  italy_2020_2021_rank <- italy %>%
    select(Season, home, visitor, hgoal, vgoal) %>%
    filter(Season == "2020" | Season == "2021") %>%
    mutate(match_outcome = case_when(
      hgoal > vgoal ~ 1, # Home team wins
      hgoal == vgoal ~ 2, # Draw
      hgoal < vgoal ~ 3 # Away team wins
    )) %>%
    mutate(periods = case_when(
      row_number() <= 190 ~ 1,
      row_number() <= 380 ~ 2,
      row_number() <= 570 ~ 3,
      TRUE ~ 4
    )) %>% # Assign periods based on match number
    select(periods,
      home_team = home,
      away_team = visitor, match_outcome
    )

  fit_rank_dyn <- btd_foot(
    data = italy_2020_2021_rank,
    dynamic_rank = TRUE,
    rank_measure = "median",
    iter_sampling = 1000,
    parallel_chains = 2,
    chains = 2
  )

  plot_logStrength(fit_rank_dyn)

  plot_logStrength(fit_rank_dyn, teams = c("AC Milan", "AS Roma", "Juventus", "Inter"))
}

## End(Not run)
```

---

pp_foot *Posterior predictive checks for football models*

Description

The function provides posterior predictive plots to check the adequacy of the Bayesian models as returned by the stan_foot function.

Usage

```
pp_foot(object, data, type = "aggregated", coverage = 0.95)
```

Arguments

object	An object either of class stanFoot, CmdStanFit, stanfit.
data	A data frame containing match data with columns: • periods: Time point of each observation (integer >= 1). • home_team: Home team's name (character string). • away_team: Away team's name (character string). • home_goals: Goals scored by the home team (integer >= 0). • away_goals: Goals scored by the away team (integer >= 0). If the model was fitted with predict > 0, the in-sample matches are used.
type	Type of plots, one among "aggregated" or "matches". Default is "aggregated".
coverage	Argument to specify the width $1 - \alpha$ of posterior probability intervals. Default is 0.95.

Details

Posterior predictive plots: when "aggregated" (default) is selected, the function returns a frequency plot for some pre-selected goal-difference values, along with their correspondent Bayesian p-values, computed as $Pr(y_{rep} \geq y)|y)$, where y_{rep} is a data replication from the posterior predictive distribution (more details in Gelman et al., 2013). Bayesian p-values very close to 0 or 1 could exhibit possible model misfits.

When "matches" is selected an ordered-frequency plot for all the goal-differences in the considered matches is provided, along with the empirical Bayesian coverage at level $1 - \alpha$.

Value

A list with elements:

- pp_plot: A ggplot object for the selected type of plot.
- pp_table: A data frame of summary statistics:
  - For "aggregated": Goal differences and their Bayesian p-values.
  - For "matches": Nominal 1-alpha and observed empirical Bayesian coverage.

Author(s)

Leonardo Egidi <legidi@units.it> and Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>

References

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis. CRC press.

Examples

```
## Not run:
if (instantiate::stan_cmdstan_exists()) {
  library(dplyr)

  data("italy")
  italy_2000 <- italy %>%
    dplyr::select(Season, home, visitor, hgoal, vgoal) %>%
    dplyr::filter(Season == "2000")

 colnames(italy_2000) <- c("periods", "home_team", "away_team", "home_goals", "away_goals")

  fit <- stan_foot(italy_2000, "double_pois", iter_sampling = 200)

  pp_foot(fit, italy_2000)
}

## End(Not run)
```

`print.btdFoot` *Print Method for btdFoot Objects*

Description

Provides detailed posterior summaries for the Bayesian Bradley-Terry-Davidson model parameters.

Usage

```
## S3 method for class 'btdFoot'
print(
  x,
  pars = NULL,
  teams = NULL,
  digits = 3,
  true_names = TRUE,
  display = "both",
  ...
)
```

Arguments

`x`	An object of class `btdFoot`.
`pars`	Optional character vector specifying parameters to include in the summary (e.g., `"logStrength"`, `"logTie"`, `"home"`, `"log_lik"`, and `"y_rep"`).
`teams`	Optional character vector specifying team names whose `logStrength` parameters should be displayed.

digits	Number of digits to use when printing numeric values. Default is 3.
true_names	Logical value indicating whether to display team names in parameter summaries. Default is TRUE.
display	Character string specifying which parts of the output to display. Options are "both", "rankings", or "parameters". Default is "both".
...	Additional arguments passed.

Author(s)

Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>

`print.compareFoot` *Print method for compareFoot objects*

Description

Provides a formatted output when printing objects of class compareFoot, displaying the predictive performance metrics and, if available, the confusion matrices for each model or probability matrix.

Usage

```
## S3 method for class 'compareFoot'
print(x, digits = 3, ...)
```

Arguments

x	An object of class compareFoot returned by compare_foot.
digits	Number of digits to use when printing numeric values for the metrics. Default is 3.
...	Additional arguments passed to print.

Author(s)

Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>

print.stanFoot *Print Method for stanFoot Objects*

Description

Provides detailed posterior summaries for the Stan football model parameters.

Usage

```
## S3 method for class 'stanFoot'
print(x, pars = NULL, teams = NULL, digits = 3, true_names = TRUE, ...)
```

Arguments

x	An object of class stanFoot.
pars	Optional character vector specifying parameters to include in the summary. This can be specific parameter names (e.g., "att", "def", "att_raw", "def_raw", "home", "sigma_att", "sigma_def", "rho", and "beta"). If NULL, all parameters are included.
teams	Optional character vector specifying team names whose "att", "def", "att_raw", "def_raw" parameters should be displayed.
digits	Number of digits to use when printing numeric values. Default is 3.
true_names	Logical value indicating whether to display team names in parameter summaries. Default is TRUE.
...	Additional arguments passed.

Author(s)

Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>

priors *Football priors distributions and options*

Description

This prior specification is just a duplicate of some of the priors used by the **rstanarm** package.

These prior distributions can be passed to the stan_foot function, through the elements ability and ability_sd of the argument prior_par. See the vignette *Prior Distributions for rstanarm Models* for further details (to view the priors used for an existing model see prior_summary). The default priors used in the **stan_foot** modeling function are intended to be *weakly informative* in that they provide moderate regularization and help stabilize computation.

You can choose between: normal, cauchy, laplace, student_t.

Usage

```
normal(location = 0, scale = NULL, autoscale = TRUE)

student_t(df = 1, location = 0, scale = NULL, autoscale = TRUE)

cauchy(location = 0, scale = NULL, autoscale = TRUE)

laplace(location = 0, scale = NULL, autoscale = TRUE)
```

Arguments

location	Prior location. In most cases, this is the prior mean, but for cauchy (which is equivalent to student_t with df=1), the mean does not exist and location is the prior median. The default value is 0.
scale	Prior scale. The default depends on the family (see **Details**).
autoscale	A logical scalar, defaulting to TRUE.
df	Prior degrees of freedom. The default is 1 for student_t, in which case it is equivalent to cauchy.

Details

The details depend on the family of the prior being used:

Student t family: Family members:

- normal(location, scale)
- student_t(df, location, scale)
- cauchy(location, scale)

Each of these functions also takes an argument autoscale.

For the prior distribution for the intercept, location, scale, and df should be scalars. For the prior for the other coefficients they can either be vectors of length equal to the number of coefficients (not including the intercept), or they can be scalars, in which case they will be recycled to the appropriate length. As the degrees of freedom approaches infinity, the Student t distribution approaches the normal distribution and if the degrees of freedom are one, then the Student t distribution is the Cauchy distribution.

If scale is not specified it will default to 10 for the intercept and 2.5 for the other coefficients.

If the autoscale argument is TRUE (the default), then the scales will be further adjusted as described above in the documentation of the autoscale argument in the **Arguments** section.

Laplace family: Family members:

- laplace(location, scale)

Each of these functions also takes an argument autoscale.

The Laplace distribution is also known as the double-exponential distribution. It is a symmetric distribution with a sharp peak at its mean / median / mode and fairly long tails. This distribution can be motivated as a scale mixture of normal distributions and the remarks above about the normal distribution apply here as well.

Value

A named list to be used internally by the `stan_foot` model fitting function.

Author(s)

Leonardo Egidi <legidi@units.it>

References

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*. 2(4), 1360–1383.

See Also

The various vignettes for the **rstanarm** package also discuss and demonstrate the use of some of the supported prior distributions.

`stan_foot` *Fit football models using CmdStan*

Description

Fits football goal-based models using Stan via the CmdStan backend. Supported models include: double Poisson, bivariate Poisson, Dixon-Coles, Skellam, Student's t, diagonal-inflated bivariate Poisson, zero-inflated Skellam, and negative Binomial.

Usage

```
stan_foot(
  data,
  model,
  predict = 0,
  ranking,
  dynamic_type,
  dynamic_weight = FALSE,
 dynamic_par = list(common_sd = FALSE, kl_variance = FALSE, periods_per_season = 2,
    spike = normal(9, 1.5), slab = normal(0, 3)),
 prior_par = list(ability = normal(0, NULL), ability_sd = cauchy(0, 5), home = normal(0,
    5)),
  home_effect = TRUE,
  norm_method = "none",
  ranking_map = NULL,
  method = "MCMC",
  ...
)
```

Arguments

data	A data frame containing match data with columns: • periods: Time point of each observation (integer >= 1). • home_team: Home team's name (character string). • away_team: Away team's name (character string). • home_goals: Goals scored by the home team (integer >= 0). • away_goals: Goals scored by the away team (integer >= 0).
model	A character string specifying the Stan model to fit. Options are: • "double_pois": Double Poisson model. • "biv_pois": Bivariate Poisson model. • "dixon_coles": Dixon-Coles model. • "neg_bin": Negative Binomial model. • "skellam": Skellam model. • "student_t": Student's t model. • "diag_infl_biv_pois": Diagonal-inflated bivariate Poisson model. • "zero_infl_skellam": Zero-inflated Skellam model.
predict	An integer specifying the number of out-of-sample matches for prediction. If missing, the function fits the model to the entire dataset without making predictions.
ranking	An optional "btdFoot" class element or a data frame containing ranking points for teams with the following columns: • periods: Time periods corresponding to the rankings (integer >= 1). • team: Team names matching those in data (character string). • rank_points: Ranking points for each team (numeric). The ranking points are matched to the teams of data by name, so the order of the rows does not matter. Every team of data must have exactly one value in each ranking period; teams of the ranking that are not in data are ignored.
dynamic_type	A character string specifying the type of dynamics in the model. Options are: • "weekly": Weekly dynamic parameters. • "seasonal": Seasonal dynamic parameters.
dynamic_weight	A logical value indicating whether to fit the weighted dynamic model of Macrì Demartino, Egidi and Torelli (2026), where the evolution of the attack and defence abilities is governed by team- and period-specific commensurate priors with spike-and-slab hyperpriors (default FALSE). It requires dynamic_type and a normal prior for the abilities (prior_par$ability = normal(...)), and it is not available for the "student_t" model. See Details.
dynamic_par	A list of options for the evolution variance of the dynamic models (ignored when dynamic_type is missing). Elements: • common_sd: A logical value. If TRUE, attack and defence abilities share a single evolution standard deviation (sigma_common), as in Owen (2011). If FALSE (default), separate standard deviations sigma_att and sigma_def are estimated, as in Egidi et al. (2018).

- kl_variance: A logical value. If TRUE, the evolution variance is inflated in the periods that follow a summer break, as in Koopman and Lit (2015). Default FALSE. It requires dynamic_type = "seasonal" and more than periods_per_season training periods.
- periods_per_season: An integer (>= 2) giving the number of consecutive values of the periods column that form one season. Used only when kl_variance = TRUE. Default 2 (each season split into two halves).
- spike: Half-normal prior normal(location, scale) for the spike component of the commensurate precisions, used when dynamic_weight = TRUE. Default normal(9, 1.5).
- slab: Half-normal prior normal(location, scale) for the slab component of the commensurate precisions, used when dynamic_weight = TRUE. Default normal(0, 3).

The options common_sd = TRUE, kl_variance = TRUE and dynamic_weight = TRUE are mutually exclusive.

prior_par — A list specifying the prior distributions for the parameters of interest:

- ability: Prior distribution for team-specific abilities. Possible distributions are normal, student_t, cauchy, laplace. Default is normal(0, NULL). In the dynamic models the prior is the distribution of the period-to-period increments of the abilities; with dynamic_weight = TRUE only normal is allowed. For student_t, cauchy and laplace the evolution standard deviations are scale parameters, not standard deviations of the increments.
- ability_sd: Prior distribution for the team-specific standard deviations. Possible distributions are normal, student_t, cauchy, laplace. Default is cauchy(0, 5).
- home: Prior distribution for the home effect (home). Applicable only if home_effect = TRUE. Only normal priors are allowed. Default is normal(0, 5).

See the **rstanarm** package for more details on specifying priors.

home_effect — A logical value indicating the inclusion of a home effect in the model. (default is TRUE).

norm_method — A character string specifying the method used to normalize team-specific ranking points. Options are:

- "none": No normalization (default).
- "standard": Standardization to mean 0 and standard deviation 0.5, i.e. the points minus their mean, divided by twice their standard deviation.
- "mad": Median Absolute Deviation normalization.
- "min_max": Min-max scaling to [0,1].

ranking_map — An optional vector mapping ranking periods to data periods. If not provided and the number of ranking periods matches the number of data periods, a direct mapping is assumed.

method — A character string specifying the method used to obtain the Bayesian estimates. Options are:

- "MCMC": Markov chain Monte Carlo algorithm (default).
- "VI": Automatic differentiation variational inference algorithms.
- "pathfinder": Pathfinder variational inference algorithm.
- "laplace": Laplace algorithm.

... Additional arguments passed to cmdstanr (e.g., iter_sampling, chains, parallel_chains).

Details

Let (y_n^H, y_n^A) denote the observed number of goals scored by the home and the away team in the n-th game, respectively. A general bivariate Poisson model allowing for goals' correlation (Karlis & Ntzoufras, 2003) is the following:

$$
\begin{aligned}
Y_n^H, Y_n^A | \lambda_{1n}, \lambda_{2n}, \lambda_{3n} &\sim \mathsf{BivPoisson}(\lambda_{1n}, \lambda_{2n}, \lambda_{3n}) \\
\log(\lambda_{1n}) &= \mu + att_{h_n} + def_{a_n} \\
\log(\lambda_{2n}) &= att_{a_n} + def_{h_n} \\
\log(\lambda_{3n}) &= \beta_0,
\end{aligned}
$$

where the case $\lambda_{3n} = 0$ reduces to the double Poisson model (Baio & Blangiardo, 2010). $\lambda_{1n}, \lambda_{2n}$ represent the scoring rates for the home and the away team, respectively, where: μ is the home effect; the parameters att_T and def_T represent the attack and the defence abilities, respectively, for each team T, $T = 1, \ldots, N_T$; the nested indexes $h_n, a_n = 1, \ldots, N_T$ denote the home and the away team playing in the n-th game, respectively. Attack/defence parameters are imposed a sum-to-zero constraint to achieve identifiability and assigned some weakly-informative prior distributions:

$$
\begin{aligned}
att_T &\sim \mathrm{N}(\mu_{att}, \sigma_{att}) \\
def_T &\sim \mathrm{N}(\mu_{def}, \sigma_{def}),
\end{aligned}
$$

with hyperparameters $\mu_{att}, \sigma_{att}, \mu_{def}, \sigma_{def}$.

The Dixon and Coles (1997) model keeps the two scoring rates $\lambda_{1n}, \lambda_{2n}$ of the double Poisson model but multiplies the joint probability of low-scoring results by a dependence factor τ to better capture the observed correlation in the 0-0, 1-0, 0-1 and 1-1 outcomes:

$$
P(Y_n^H = y_n^H, Y_n^A = y_n^A) = \tau_{\lambda_{1n}, \lambda_{2n}}(y_n^H, y_n^A)\, \mathsf{Poisson}(y_n^H | \lambda_{1n})\, \mathsf{Poisson}(y_n^A | \lambda_{2n}),
$$

where

$$
\tau_{\lambda_1, \lambda_2}(y^H, y^A) = \begin{cases} 1 - \lambda_1 \lambda_2 \rho & y^H = y^A = 0 \\ 1 + \lambda_1 \rho & y^H = 0, y^A = 1 \\ 1 + \lambda_2 \rho & y^H = 1, y^A = 0 \\ 1 - \rho & y^H = y^A = 1 \\ 1 & \text{otherwise,} \end{cases}
$$

and ρ is the low-score dependence parameter, assigned a weakly-informative $\mathrm{N}(0, 0.1)$ prior and constrained to keep the adjustment factor positive.

Instead of using the marginal number of goals, another alternative is to modelling directly the score difference ($y_n^H - y_n^A$). We can use the Poisson-difference distribution (or Skellam distribution) to model goal difference in the n-th match (Karlis & Ntzoufras, 2009):

$$y_n^H - y_n^A | \lambda_{1n}, \lambda_{2n} \sim PD(\lambda_{1n}, \lambda_{2n}),$$

and the scoring rates $\lambda_{1n}, \lambda_{2n}$ are unchanged with respect to the bivariate/double Poisson model. If we want to use a continue distribution, we can use a student t distribution with 7 degrees of freedom (Gelman, 2014):

$$y_n^H - y_n^A \sim t(7, ab_{h_n} - ab_{a(n)}, \sigma_y)$$
$$ab_t \sim \mathrm{N}(\mu + b \times prior_score_t, sigma_{ab}),$$

where ab_t is the overall ability for the t-th team, whereas $prior_score_t$ is a prior measure of team's strength (for instance a ranking).

These model rely on the assumption of static parameters. However, we could assume dynamics in the attach/defence abilities (Owen, 2011; Egidi et al., 2018, Macri Demartino et al., 2024) in terms of weeks or seasons through the argument `dynamic_type`. In such a framework, for a given number of times $1, \ldots, \mathcal{T}$, the models above would be unchanged, but the priors for the abilities parameters at each time $\tau, \tau = 2, \ldots, \mathcal{T}$, would be:

$$att_{T,\tau} \sim \mathrm{N}(att_{T,\tau-1}, \sigma_{att})$$
$$def_{T,\tau} \sim \mathrm{N}(def_{T,\tau-1}, \sigma_{def}),$$

whereas for $\tau = 1$ we have:

$$att_{T,1} \sim \mathrm{N}(\mu_{att}, \sigma_{att})$$
$$def_{T,1} \sim \mathrm{N}(\mu_{def}, \sigma_{def}).$$

Of course, the identifiability constraint must be imposed for each time τ.

The evolution standard deviations of the dynamic models can be specified in four alternative ways, through the arguments `dynamic_par` and `dynamic_weight`:

- Default (Egidi et al., 2018): two distinct standard deviations σ_{att} and σ_{def}, shared by all the teams and periods (parameters `sigma_att` and `sigma_def`).
- `dynamic_par = list(common_sd = TRUE)` (Owen, 2011): a single evolution standard deviation $\sigma_{att} = \sigma_{def} = \sigma$ (parameter `sigma_common`).
- `dynamic_par = list(kl_variance = TRUE)` (Koopman and Lit, 2015): the evolution variance is inflated in the periods that follow a summer break,

  $$\sigma_{k,\tau}^2 = \sigma_k^2 + \sigma_{break}^2 I_\tau, \quad k \in \{att, def\},$$

  where $I_\tau = 1$ if period τ follows a summer break and 0 otherwise. The package assumes that each season is made of `periods_per_season` consecutive periods (default 2, i.e. two half-seasons), so that a summer break precedes the periods $1 + m\,\text{periods_per_season}$, $m = 1, 2, \ldots$ (parameters `sigma_att_kl`, `sigma_def_kl` and `sigma_break`).

- `dynamic_weight = TRUE` (Macrì Demartino et al., 2026): weighted dynamic model with commensurate priors. For each team T, ability $k \in \{att, def\}$ and period $\tau \geq 2$,

$$k_{T,\tau} \sim \mathrm{N}(k_{T,\tau-1}, 1/\phi_{k,T,\tau}),$$

where the commensurate precision $\phi_{k,T,\tau}$ is assigned a two-component mixture of half-normal distributions (spike-and-slab),

$$\phi_{k,T,\tau} \sim p_T\, \mathrm{N}^{+}(\mu_s, \psi_s) + (1 - p_T)\, \mathrm{N}^{+}(\mu_l, \psi_l), \quad p_T \sim \mathrm{Beta}(1, 1).$$

The spike $\mathrm{N}^{+}(\mu_s, \psi_s)$ is concentrated on large precisions (strong borrowing of information from the previous period), whereas the slab $\mathrm{N}^{+}(\mu_l, \psi_l)$ is diffuse on small precisions (weak borrowing). The hyperparameters are set through `dynamic_par$spike` and `dynamic_par$slab` (parameters `prob_spike`, `comm_prec_att`, `comm_prec_def`, `comm_sd_att` and `comm_sd_def`).

To better capture draw occurrences, the package also allows for the fit of a diagonal-inflated bivariate Poisson and a zero-inflated Skellam model. The diagonal-inflated bivariate Poisson model (Karlis & Ntzoufras, 2003) is a mixture of a bivariate Poisson and a distribution over the draws:

$$P(Y_n^H = x, Y_n^A = y) = (1 - p)\, \mathsf{BivPoisson}(x, y | \lambda_{1n}, \lambda_{2n}, \lambda_{3n}) + p\, \theta_x\, I(x = y),$$

where p is the probability of the inflation component (parameter `prob_of_draws`, with a $\mathrm{U}(0, 1)$ prior) and θ_j, $j = 0, \ldots, 3$, is a discrete distribution over the draws 0-0, 1-1, 2-2 and 3-3 (parameter `draw_dist`, with a uniform Dirichlet prior); draws with more than three goals per team are not inflated. See the vignette for further details.

Value

An object of class `"stanFoot"`, which is a list containing:

- `fit`: The `CmdStanFit` object returned by `cmdstanr`.
- `data`: The input data.
- `stan_data`: The data list passed to Stan.
- `stan_code`: The Stan code of the underline model.
- `stan_args`: The optional `cmdstanr` parameters passed to (`...`).
- `alg_method`: The inference algorithm used to obtain the Bayesian estimates.

Author(s)

Leonardo Egidi <legidi@units.it>, Roberto Macrì Demartino <roberto.macridemartino@deams.units.it>, and Vasilis Palaskas <vasilis.palaskas94@gmail.com>.

References

Baio, G. and Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. Journal of Applied Statistics 37(2), 253-264.

Dixon, M. J. and Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. Journal of the Royal Statistical Society: Series C (Applied Statistics), 46(2), 265-280.

Egidi, L., Pauli, F., and Torelli, N. (2018). Combining historical data and bookmakers' odds in modelling football scores. Statistical Modelling, 18(5-6), 436-459.

Gelman, A. (2014). Stan goes to the World Cup. From "Statistical Modeling, Causal Inference, and Social Science" blog.

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J. and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics, 67(3), 1047-1056.

Koopman, S. J. and Lit, R. (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. Journal of the Royal Statistical Society: Series A (Statistics in Society), 178(1), 167-186.

Macrì Demartino, R., Egidi, L. and Torelli, N. (2024). Alternative ranking measures to predict international football results. Computational Statistics, 1-19.

Macrì Demartino, R., Egidi, L. and Torelli, N. (2026). Bayesian weighted discrete-time dynamic models for association football prediction. Journal of the Royal Statistical Society Series C: Applied Statistics, qlag032. doi:10.1093/jrsssc/qlag032

Karlis, D. and Ntzoufras, I. (2003). Analysis of sports data by using bivariate poisson models. Journal of the Royal Statistical Society: Series D (The Statistician) 52(3), 381-393.

Karlis, D. and Ntzoufras,I. (2009). Bayesian modelling of football outcomes: Using the Skellam's distribution for the goal difference. IMA Journal of Management Mathematics 20(2), 133-145.

Owen, A. (2011). Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. IMA Journal of Management Mathematics, 22(2), 99-113.

Examples

```
## Not run:
if (instantiate::stan_cmdstan_exists()) {
  library(dplyr)

  data("italy")
  italy <- as_tibble(italy)

  # Four seasons of the Italian Serie A, each split into two halves
  # (periods 1-8): the summer break precedes periods 3, 5 and 7
  italy_2018_2021 <- italy %>%
    select(Season, home, visitor, hgoal, vgoal) %>%
    filter(Season %in% c("2018", "2019", "2020", "2021")) %>%
    group_by(Season) %>%
    mutate(half = if_else(row_number() <= n() / 2, 1, 2)) %>%
    ungroup() %>%
    mutate(periods = 2 * (as.numeric(Season) - 2018) + half) %>%
    select(periods,
      home_team = home, away_team = visitor,
      home_goals = hgoal, away_goals = vgoal
    )

  # Koopman & Lit (2015): variance inflation at the summer breaks
  fit_kl <- stan_foot(
    data = italy_2018_2021,
```

```
  model = "biv_pois",
  dynamic_type = "seasonal",
  dynamic_par = list(kl_variance = TRUE, periods_per_season = 2),
  iter_sampling = 1000,
  chains = 4,
  parallel_chains = 4
)
print(fit_kl, pars = c("sigma_att_kl", "sigma_def_kl", "sigma_break"))

# Weighted dynamic model with commensurate priors
# (Macrì Demartino et al., 2026)
fit_wdm <- stan_foot(
  data = italy_2018_2021,
  model = "double_pois",
  dynamic_type = "seasonal",
  dynamic_weight = TRUE,
  dynamic_par = list(spike = normal(9, 1.5), slab = normal(0, 3)),
  iter_sampling = 1000,
  chains = 4,
  parallel_chains = 4
)
print(fit_wdm, pars = "prob_spike")
}

## End(Not run)
```

Index

∗ **datasets**
    england, 8
    italy, 16

btd_foot, 3

cauchy *(*priors*)*, 27
cmdstanr, *4*, *32*, *34*
compare_foot, 6, *26*

england, 8

foot_abilities, 9
foot_prob, 11
foot_rank, 13
foot_round_robin, 14

italy, 16

laplace *(*priors*)*, 27
list, *9*, *12*

mle_foot, 17

normal *(*priors*)*, 27

plot_btdPosterior, 20
plot_logStrength, 22
pp_foot, 23
print.btdFoot, 25
print.compareFoot, 26
print.stanFoot, 27
priors, 27

stan_foot, 29
stanfit, *12*
stats, *18*
student_t *(*priors*)*, 27