Package 'lddmm'

October 23, 2026

Type Package

Title Longitudinal Drift-Diffusion Mixed Models (LDDMM)

Version 1.0.0

Date 2026-10-22

Description Implementation of the drift-diffusion mixed model for category learning as described in Paulon et al. (2021) <doi:10.1080/01621459.2020.1801448>.

Depends R (>= 4.1.0)

Language en-US

License MIT + file LICENSE

URL https://github.com/giorgiopaulon/LDDMM

BugReports https://github.com/giorgiopaulon/LDDMM/issues

Encoding UTF-8

Imports Rcpp (>= 1.0.6), dplyr, tidyr, ggplot2 (>= 3.4.0), latex2exp, reshape2, stats, utils

LazyData true

LinkingTo Rcpp, RcppArmadillo

Suggests rmarkdown, knitr, testthat (>= 3.0.0), statmod

Config/testthat/edition 3

VignetteBuilder knitr

Config/roxygen2/version 8.1.1

NeedsCompilation yes

Author Giorgio Paulon [aut, cre], Abhra Sarkar [aut, ctb]

Maintainer Giorgio Paulon <giorgiopaulon90@gmail.com>

Repository CRAN

Date/Publication 2026-10-23 09:10:34 UTC

Contents

B_basis . . . 2
compute_WAIC . . . 3
data . . . 3
extract_post_draws . . . 4
extract_post_mean . . . 4
H_ball . . . 5
LDDMM . . . 5
log_likelihood . . . 8
log_likelihood_ind . . . 9
plot_accuracy . . . 9
plot_post_pars . . . 10
plot_predictive . . . 10
plot_RT . . . 11
P_smooth1 . . . 12

Index **13**

`B_basis` *Spline Basis Functions*

Description

Quadratic B-spline basis on equally spaced knots (supplement §S.1), evaluated on a grid. Points outside `[min(knots), max(knots)]` get a row of zeros.

Usage

```
B_basis(xgrid, knots)
```

Arguments

`xgrid`	grid where we want to evaluate the spline functions (vector of length n)
`knots`	vector of equally spaced knots for the splines (vector of length K)

Value

n x (K+1) matrix with the value of each basis function on `xgrid`. At a knot, exactly two adjacent basis functions are non-zero, both equal to 1/2.

compute_WAIC *Calculate WAIC*

Description

Watanabe-Akaike information criterion (Gelman, Hwang and Vehtari, 2014), which estimates the expected out-of-sample prediction error using a bias-corrected adjustment of the within-sample error.

Usage

```
compute_WAIC(model_fit)
```

Arguments

model_fit	results of a model fit from the LDDMM() function, run with save_loglik = TRUE

Value

A scalar indicating the WAIC (smaller WAIC denotes better fit)

data *Example dataset*

Description

Data in the format required by LDDMM(): 24,254 trials of 99 participants over 5 training blocks, with two possible input stimuli and decisions. The data were analyzed in Roark, Paulon, Sarkar and Chandrasekaran (2021), Psychonomic Bulletin & Review 28, 898-909.

Usage

```
data
```

Format

A data frame with 24,254 rows and 6 columns.

Details

- subject: participant labels
- block: training blocks (longitudinal units)
- s: input stimuli
- d: decisions
- r_time: response times, in seconds
- cens: censoring indicators (1 censored, 0 non censored); all 0 here

extract_post_draws *Parameter posterior draws*

Description

Posterior draws of the subject-level drift or boundary parameters of the correct responses (d = s) from an LDDMM() fit.

Usage

```
extract_post_draws(data, fit, par = c("drift", "boundary"))
```

Arguments

data	the data frame used to fit the model (see LDDMM()).
fit	fit from the LDDMM() function
par	parameter to output ('drift', or 'boundary')

Value

Data frame with columns block, subject, draw and one column <par>_<s>_<s> per stimulus with the posterior draws.

extract_post_mean *Parameter posterior means*

Description

Posterior means of the subject-level drift or boundary parameters of the correct responses (d = s) from an LDDMM() fit.

Usage

```
extract_post_mean(data, fit, par = c("drift", "boundary"))
```

Arguments

data	the data frame used to fit the model (see LDDMM()).
fit	fit from the LDDMM() function
par	parameter to output ('drift', or 'boundary')

Value

Data frame with columns block, subject and one column <par>_<s>_<s> per stimulus with the posterior means.

H_ball *Hamming Ball*

Description

All label vectors within Hamming distance `r` of `x` (supplement §S.4).

Usage

```
H_ball(x, S, r)
```

Arguments

`x`	center of the Hamming Ball (vector of labels in `1:S`)
`S`	number of states
`r`	radius of the Hamming Ball

Value

Matrix with `length(x)` rows; each column is one member of the ball, starting with `x` itself and ordered by distance from `x`.

LDDMM *Drift Diffusion Model Fit*

Description

Fits the longitudinal inverse Gaussian drift-diffusion mixed model of Paulon et al. (2021) by Markov chain Monte Carlo.

Usage

```
LDDMM(
  data,
  hypers = NULL,
  cluster = TRUE,
  boundaries = c("flexible", "constant", "fixed", "fixed-constant"),
  random_effects = c("correct-incorrect", "shared"),
  z_max = NULL,
  informed = TRUE,
  Niter = 5000,
  burnin = 2000,
  thin = 5,
  save_z = FALSE,
  save_loglik = TRUE,
```

```
  verbose = TRUE,
  debug_checks = FALSE
)
```

Arguments

`data`	data frame with one row per trial and the columns • `subject`: participant labels • `block`: training blocks (longitudinal units), consecutive integers • `s`: input stimuli, coded 1, ..., K • `d`: decisions, coded 1, ..., K like `s` (may be `NA` for censored trials) • `r_time`: response times (positive), e.g. in seconds • `cens`: censoring indicators (1 censored, 0 observed); for a censored trial `r_time` is the time at which observation stopped
`hypers`	list of hyperparameters (missing entries take default values): • `s_sigma_mu`, `s_sigma_b`: scales of the half-Cauchy priors on the standard deviations σ_1 that control the smoothness of the drift and boundary curves (default 0.1) • `sigma2_1_max`, `sigma2_us_max`: upper truncation of the priors on the smoothness variances of the fixed-effect curves and of the random-effect curves (default 0.2) • `sigma2_ua_max`: upper truncation of the half-Cauchy prior on the variance of the random effects (default `Inf`, no truncation) • `pseudo_sd`: spread of the pseudo-prior of unused clusters (default 0.1). Coefficients of clusters that no curve uses are drawn close to those of the clusters in use, so that curves can split off into a new cluster. This only affects how well the sampler mixes, not the posterior. • `mu_beta0_mu`, `sigma2_beta0_mu`, `mu_beta0_b`, `sigma2_beta0_b`: mean and variance of the normal prior on spline coefficients that start a curve or belong to an unused cluster (article eq. 9). By default they are set from moment estimates in every (block, stimulus, decision) cell.
`cluster`	should the fixed-effect coefficients be locally clustered across input-response combinations? With `FALSE` every combination has its own curve.
`boundaries`	boundary model: • `"flexible"`: $b_{d,s}(t)$ varies over time and is clustered jointly with the drifts • `"fixed"`: $b_d(t)$ varies over time but not across stimuli • `"constant"`: $b_{d,s}$ is constant over time • `"fixed-constant"`: b_d is constant over time and across stimuli The random effects of the boundary are spline curves for the first two models and scalars for the last two.
`random_effects`	`"correct-incorrect"` gives every subject one random-effect curve for correct responses ($d = s$) and another for incorrect responses (article §3.2.2); `"shared"` uses a single curve for both.
`z_max`	maximum number of clusters at each knot when `cluster = TRUE`. Defaults to `min(K^2, 2K)`, i.e. 8 for 4 categories as in the article.

`informed`	use locally informed proposals in the Hamming ball sampler of the cluster labels (supplement §S.4)?
`Niter`	total number of iterations
`burnin`	burnin of the chain; proposal scales are tuned only during burn-in
`thin`	thinning factor
`save_z`	store the cluster labels of every saved draw (element `Z`)?
`save_loglik`	store the log-likelihood of every observation for every saved draw (element `loglik`, needed by `compute_WAIC()`)?
`verbose`	show a progress bar?
`debug_checks`	verify after every update that the sampler's cached quantities match their definitions (slow; for testing).

Details

On trial ℓ of block t, subject i hears input stimulus s and answers d. One accumulator per possible decision races to its boundary; the first to arrive determines the decision and the response time. The accumulator of decision d has drift $\mu_{d,s}^{(i)}(t) = \exp\{f_{\mu,d,s}(t) + u_{\mu}^{(i)}(t)\}$, boundary $b_{d,s}^{(i)}(t) = \exp\{f_{b,d,s}(t) + u_{b}^{(i)}(t)\}$ and starts after an offset $\delta_s^{(i)}$. The fixed effects f are quadratic B-splines over blocks whose coefficients are locally clustered across (d, s) by hidden Markov models; the random effects u are smooth subject-specific spline curves.

Priors are calibrated so that, for the most stable performance, most values of `r_time` should lie between 0 and 10; response times in seconds work well. The offset $\delta_s^{(i)}$ cannot exceed the fastest response of subject i to stimulus s, so implausibly fast responses (for example anticipations under 100 ms) force it towards zero and are best removed before fitting.

Value

List with the following MCMC posterior samples:

- `post_mean_delta`: population offsets (draw x stimulus)
- `post_mean_mu`, `post_mean_b`: population drift and boundary curves (block x stimulus x decision x draw), integrating out the random effects
- `post_ind_delta`: individual offsets (stimulus x subject x draw)
- `post_ind_mu`, `post_ind_b`: individual drift and boundary curves (block x subject x stimulus x decision x draw)
- `sigma2_mu_us`, `sigma2_mu_ua`, `sigma2_b_us`, `sigma2_b_ua`: smoothness and variance parameters of the random effects
- `sigma2_1_mu`, `sigma2_1_b`: smoothness parameters of the fixed effects (`sigma2_b_us` and `sigma2_1_b` are `NA` for constant boundaries)
- `pred_ans`, `pred_time`: posterior predictive decisions and response times at the population level (block x stimulus x draw)
- `pred_ans_ind`, `pred_time_ind`: the same for each subject (subject x block x stimulus x draw)
- `loglik`: log-likelihood of each observation (draw x observation)

- Z: cluster labels (stimulus x decision x knot x draw), if save_z = TRUE
- acceptance: acceptance rates of the Metropolis-Hastings steps after burn-in
- settings: model options and hyperparameters used

References

Paulon, G., Llanos, F., Chandrasekaran, B., Sarkar, A. (2021). Bayesian semiparametric longitudinal drift-diffusion mixed models for tone learning in adults. Journal of the American Statistical Association 116, 1114-1127. doi:10.1080/01621459.2020.1801448

`log_likelihood` *Log-likelihood*

Description

Log-likelihood of the inverse Gaussian race model (article eq. 2), including the contribution of censored observations.

Usage

```
log_likelihood(tau, mu, b, delta, cens, D, log)
```

Arguments

tau	vector of size n containing the response times
mu	matrix of size (n x d1) containing the drift parameters corresponding to the n response times for each possible d1 decision
b	matrix of size (n x d1) containing the boundary parameters corresponding to the n response times for each possible d1 decision
delta	vector of size n containing the offset parameters corresponding to the n response times
cens	vector of size n containing censoring indicators (1 censored, 0 not censored) corresponding to the n response times
D	(n x 2) matrix whose first column has the n input stimuli, and whose second column has the n decision categories
log	should the result be returned on the log scale?

Value

The (log-)likelihood of all n observations.

log_likelihood_ind *Log-likelihood of each observation*

Description

Log-likelihood of the inverse Gaussian race model for every observation, including the contribution of censored observations.

Usage

```
log_likelihood_ind(tau, mu, b, delta, cens, D)
```

Arguments

tau	vector of size n containing the response times
mu	matrix of size (n x d1) containing the drift parameters corresponding to the n response times for each possible d1 decision
b	matrix of size (n x d1) containing the boundary parameters corresponding to the n response times for each possible d1 decision
delta	vector of size n containing the offset parameters corresponding to the n response times
cens	vector of size n containing censoring indicators (1 censored, 0 not censored) corresponding to the n response times
D	(n x 2) matrix whose first column has the n input stimuli, and whose second column has the n decision categories

Value

Vector of n log-likelihood values. An uncensored response time that is not larger than its offset has log-likelihood `-Inf`.

plot_accuracy *Descriptive plots*

Description

Proportion of each decision for every input stimulus and block, by subject (thin lines) and pooled over subjects (thick lines). Censored trials are excluded.

Usage

```
plot_accuracy(data)
```

Arguments

data	data frame in the format described in LDDMM().

Value

A ggplot object.

---

plot_post_pars *Plot posterior estimates*

Description

Posterior means and 90% credible intervals of the population-level drift or boundary curves from an LDDMM() fit. Each panel is an input stimulus s; colours are the decisions d.

Usage

```
plot_post_pars(data, fit, par = c("drift", "boundary"))
```

Arguments

data	the data frame used to fit the model (see LDDMM()).
fit	fit from the LDDMM() function
par	parameter to output ('drift', or 'boundary')

Value

A ggplot object.

---

plot_predictive *Posterior predictive checks*

Description

Compares the data with responses simulated from the fitted model. For every saved draw, LDDMM() simulates one response of every subject to every stimulus in every block (pred_ans_ind and pred_time_ind).

- type = "response": proportion of each decision, by stimulus and block. Lines are the posterior predictive mean of the proportion; bands are the 90% interval of the proportion among the simulated responses of one draw (one per participant), so they also contain the sampling variability of that many responses. Points are the observed proportions.
- type = "rt": 10%, 50% and 90% quantiles of the response times, by stimulus and block. Lines are the posterior predictive quantiles (pooled over subjects and draws); points are the observed quantiles.

Usage

```
plot_predictive(data, fit, type = c("response", "rt"))
```

Arguments

data	the data frame used to fit the model (see LDDMM()).
fit	fit from the LDDMM() function
type	which check to plot: "response" or "rt".

Details

Censored trials are left out of the observed data.

Value

A ggplot object.

---

plot_RT *Descriptive plots*

Description

Mean response time of each decision for every input stimulus and block, pooled over subjects. Censored trials are excluded.

Usage

```
plot_RT(data)
```

Arguments

data	data frame in the format described in LDDMM().

Value

A ggplot object.

P_smooth1 *Spline Penalty Matrix*

Description

Penalty matrix of the smoothness inducing prior for spline coefficients: `t(beta) %*% P %*% beta` is the sum of squared first differences of `beta` (Eilers and Marx, 1996; article eq. 10).

Usage

```
P_smooth1(K)
```

Arguments

`K`	number of spline coefficients

Value

K x K matrix `P = t(D) %*% D`, with `D` the (K-1) x K first-difference matrix.

Index

∗ **datasets**
data, 3

B_basis, 2

compute_WAIC, 3
compute_WAIC(), *7*

data, 3

extract_post_draws, 4
extract_post_mean, 4

H_ball, 5

LDDMM, 5
LDDMM(), *3, 4*, *10, 11*
log_likelihood, 8
log_likelihood_ind, 9

P_smooth1, 12
plot_accuracy, 9
plot_post_pars, 10
plot_predictive, 10
plot_RT, 11