

Package ‘fddm’

November 4, 2020

Type Package

Title Fast Implementation of the Diffusion Decision Model

Version 0.2-2

Description Provides the probability density function (PDF) of the diffusion decision model (DDM; e.g., Ratcliff & McKoon, 2008, <doi:10.1162/neco.2008.12-06-420>) with across-trial variability in the drift rate. Because the PDF of the DDM contains an infinite sum, it needs to be approximated. 'fddm' implements all published approximations (Navarro & Fuss, 2009, <doi:10.1016/j.jmp.2009.02.003>; Gondan, Blurton, & Kesselmeier, 2014, <doi:10.1016/j.jmp.2014.05.002>) plus new approximations. All approximations are implemented purely in 'C++' providing faster speed than existing packages.

License GPL (>= 2)

Imports Rcpp (>= 1.0.1)

LinkingTo Rcpp

Depends R (>= 3.5.0)

Suggests rtdists, RWiener, ggplot2, reshape2, testthat, knitr, rmarkdown, microbenchmark, ggnewscale

URL <https://github.com/rtdists/fddm>

BugReports <https://github.com/rtdists/fddm/issues>

SystemRequirements C++11

RoxygenNote 7.1.1

Encoding UTF-8

VignetteBuilder knitr

NeedsCompilation yes

Author Kendal B. Foster [aut],
Henrik Singmann [ctb, cre] (<<https://orcid.org/0000-0002-4842-3657>>)

Maintainer Henrik Singmann <singmann@gmail.com>

Repository CRAN

Date/Publication 2020-11-04 20:00:02 UTC

R topics documented:

dfddm	2
med_dec	6

Index	8
--------------	----------

dfddm	<i>Density of Ratcliff Diffusion Decision Model</i>
-------	---

Description

Density function for the Ratcliff diffusion decision model (DDM) with following parameters: a (threshold separation), v (drift rate), t_0 (non-decision time/response time constant), w (relative starting point), and sv (inter-trial variability of drift).

Usage

```
dfddm(
  rt,
  response,
  a,
  v,
  t0,
  w = 0.5,
  sv = -1,
  log = FALSE,
  n_terms_small = "SWSE",
  summation_small = "2017",
  scale = "both",
  max_terms_large = 4,
  err_tol = 1e-06
)
```

Arguments

<code>rt</code>	A vector of response times (in seconds). If a response time is non-positive, then its density will evaluate to 0 if <code>log = FALSE</code> and $-\infty$ if <code>log = TRUE</code> .
<code>response</code>	Binary response(s) that correspond(s) to either the "lower" or "upper" threshold. This model parameter can either be a singular value or a vector. The value(s) in 'response' can be of the following data types: <ul style="list-style-type: none"> • logicals (FALSE \rightarrow "lower", TRUE \rightarrow "upper"); • integers or doubles (1 \rightarrow "lower", 2 \rightarrow "upper"); • strings (only the first character is checked, "L" \rightarrow "lower" or "U" \rightarrow "upper", case insensitive); • factors (the first level gets mapped to "lower", and the second level gets mapped to "upper"; any additional levels are ignored).

a	Threshold separation. Amount of information that is considered for a decision. Large values indicate a conservative decisional style. Allowed range: $0 < a$. Typical range: $0.5 < a < 2$.
v	Drift rate. Average slope of the information accumulation process. The drift gives information about the speed and direction of the accumulation of information. Large (absolute) values of drift indicate a good performance. If received information supports the response linked to the upper threshold, then the sign will be positive; similarly a negative value indicates that the received information supports the response linked to the lower threshold. Allowed range: v is a real number. Typical range: $-5 < v < 5$.
t0	Non-decision time or response time constant (in seconds). Lower bound for the duration of all non-decisional processes (encoding and response execution). If this value is greater than rt, then the resulting density is returned as if $rt \leq 0$. Allowed range: $0 \leq t0$. Typical range: $0.1 < t0 < 0.5$.
w	Relative starting point. Indicator of an a priori bias in decision making. When the relative starting point w deviates from 0.5, the amount of information necessary for a decision differs between response alternatives. Allowed range: $0 < w < 1$. Default is 0.5 (i.e., no bias).
sv	Inter-trial-variability of drift rate. Standard deviation of a normal distribution with mean v describing the distribution of actual drift rates from specific trials. Values different from 0 can predict slow errors. Allowed range: $0 < sv$. Typical range: $0 < sv < 2$. Default is -1, which indicates no drift in the function call. See Details for more information.
log	Logical; if TRUE, probabilities p are given as $\log(p)$. Default is FALSE.
n_terms_small	Which method for calculating number of terms used in the approximation of the infinite sum in the small-time approximation of the density function. Can be one of {"SWSE", "Gondan", "Navarro"}. Only applicable if scale is one of {"small", "both"}. See Details for more information. Default is "Gondan".
summation_small	Which style of summation to use for the small-time approximation to the infinite sum. Can be one of {"2017", "2014"}. Only applicable if scale is one of {"small", "both"}. See Details for more information. Default is "2017".
scale	Which density function to use. Can be one of {"small", "large", "both"}. Note that the large-time approximation is unstable for small effective response times ($rt/(a * a) < 0.009$). See Details for more information. Default is "both".
max_terms_large	Maximum number of terms to use for the "large-time" variant when n_terms_small = "SWSE", scale = "both". Allowed values are any non-negative integer. max_terms_large = 0 indicates that the "small-time" variant will always be used instead of the "large-time" variant. Default is 4.
err_tol	Allowed error tolerance of the density function. Since the density function contains an infinite sum, this parameter defines the precision of the approximation to that infinite sum. Default is $1e - 6$.

Details

All of the model inputs and parameters (rt , $response$, a , v , $t\theta$, w , and sv) can be input as a single value or as a vector of values. If input as a vector of values, then the standard R input wrapping will occur.

The default settings of $n_terms_small = "SWSE"$, $codesummation_small = "2017"$, $scale = "both"$ produce the fastest and most accurate results, as shown in our associated paper.

$scale$ - The density function for the DDM has traditionally been written in two forms: a "large-time" variant, and a "small-time" variant. The parameter $scale$ determines which of these variants will be used in the calculation; $scale = "large"$ uses the "large-time" variant, and $scale = "small"$ uses the "small-time" variant. The "large-time" variant is unstable for small effective response times ($rt/(a*a) < 0.009$) and produces inaccurate densities; thus we do not recommend using only the $scale = "large"$ option if the inputs contain such small response times. To circumvent this issue, the $scale = "both"$ option utilizes both the "small-time" and "large-time" variants by determining which variant is more computationally efficient before calculating the density. Even though the "large-time" density function is often significantly slower than the "small-time" variant, it is extremely efficient in some areas of the parameter space, and so the $scale = "both"$ option is the fastest.

sv - Both the "small-time" and "large-time" variants of the density function have two further variants: one with a constant drift rate v (i.e., $sv = 0$), and one with a variable drift rate v (i.e., $sv > 0$). The details of the differences between these two density functions can be found in our associated paper. To use the density function with a constant drift rate, leave the parameter sv to its default value of $sv = -1$, as this will indicate no drift to the function. To use the density function with a variable drift rate, set the parameter sv to some positive value, e.g. $sv = .75$.

$summation_small$ - The "large-time" variant of the density function does not have any further variants, but the "small-time" has more options with respect to evaluating the infinite sum. There are two equivalent styles of summation, $summation_small = "2017"$ and $summation_small = "2014"$, of which the "2017" version evaluates slightly faster and thus earns our recommendation. These different styles of summation are discussed in our associated paper.

n_terms_small - The "small-time" variant also has three different methods for how to truncate the infinite sum in the density function. These different methods are discussed extensively in our associated paper, but the key distinction is that $n_terms_small = "SWSE"$ uses a new method of truncating the infinite sum. The $n_terms_small = "SWSE"$ method is currently recommended because it is the most stable algorithm when used with $scale = "both"$.

Value

A vector containing the densities of the DDM with precision err_tol whose length matches that of the longest input parameter (usually rt).

References

- Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. *Journal of Mathematical Psychology*, 53(4), 222-230.
- Gondan, M., Blurton, S. P., & Kesselmeier, M. (2014). Even faster and even more accurate first-passage time densities and distributions for the Wiener diffusion model. *Journal of Mathematical Psychology*, 60, 20-22.

Blurton, S. P., Kesselmeier, M., & Gondan, M. (2017). The first-passage time distribution for the diffusion model with variable drift. *Journal of Mathematical Psychology*, 76, 7-12.

Examples

```
# minimal example
dfddm(rt = 1, response = "lower", a = 1, v = -2, t0 = 0.3)

# more practical example
dfddm(rt = c(3, 2.75, 1.4, 2, 2.3, 1.8),
      response = c("lower", "upper", "upper", "lower", "upper", "lower"),
      a = 1, v = -2, t0 = 0.3, w = 0.4, sv = 0.1,
      log = TRUE, n_terms_small = "SWSE",
      summation_small = "2017", scale = "both",
      err_tol = 1e-10)

# example of mismatched input lengths
dfddm(rt = c(3, 2.75, 1.4, 2, 2.3, 1.8),
      response = c("lower", "upper", "upper", "lower", "upper", "lower"),
      a = c(1, 3), v = c(-2, 2, 2, -2, 2, -2),
      t0 = 0.3, w = c(0.4, 0.6, 0.5), sv = 0.1,
      log = FALSE, n_terms_small = "Navarro",
      summation_small = "2017", scale = "large",
      err_tol = 1e-10)

### examples of different response inputs

# string
resp_str <- c("lower", "upper", "upper", "lower", "upper", "lower")
dfddm(rt = c(3, 2.75, 1.4, 2, 2.3, 1.8), response = resp_str,
      a = 1, v = -2, t0 = 0.3, w = 0.4, sv = 0.1,
      log = FALSE, n_terms_small = "SWSE",
      summation_small = "2017", scale = "both",
      err_tol = 1e-10)

# integer
resp_int <- as.integer(c(1, 2, 2, 1, 2, 1))
dfddm(rt = c(3, 2.75, 1.4, 2, 2.3, 1.8), response = resp_int,
      a = 1, v = -2, t0 = 0.3, w = 0.4, sv = 0.1,
      log = FALSE, n_terms_small = "SWSE",
      summation_small = "2017", scale = "both",
      err_tol = 1e-10)

# double
resp_dbl <- as.double(c(1, 2, 2, 1, 2, 1))
dfddm(rt = c(3, 2.75, 1.4, 2, 2.3, 1.8), response = resp_dbl,
      a = 1, v = -2, t0 = 0.3, w = 0.4, sv = 0.1,
      log = FALSE, n_terms_small = "SWSE",
      summation_small = "2017", scale = "both",
      err_tol = 1e-10)

# logical
```

```

resp_log <- c(FALSE, TRUE, TRUE, FALSE, TRUE, FALSE)
dfddm(rt = c(3, 2.75, 1.4, 2, 2.3, 1.8), response = resp_log,
      a = 1, v = -2, t0 = 0.3, w = 0.4, sv = 0.1,
      log = FALSE, n_terms_small = "SWSE",
      summation_small = "2017", scale = "both",
      err_tol = 1e-10)

# factor (first level is mapped to "lower")
days <- c("Monday", "Friday", "Friday", "Monday", "Friday", "Monday")
resp_fac <- factor(days, levels = c("Monday", "Friday"))
dfddm(rt = c(3, 2.75, 1.4, 2, 2.3, 1.8), response = resp_fac,
      a = 1, v = -2, t0 = 0.3, w = 0.4, sv = 0.1,
      log = FALSE, n_terms_small = "SWSE",
      summation_small = "2017", scale = "both",
      err_tol = 1e-10)

```

med_dec

Medicial decision data

Description

Part of the accuracy and response time data presented in Trueblood et al. (2017) investigating medical decision making among medical professionals (pathologists) and novices (i.e., undergraduate students). The task of participants was to judge whether pictures of blood cells show cancerous cells (i.e., blast cells) or non-cancerous cells (i.e., non-blast cells). The current data set contains 200 decisions per participant (the "accuracy" condition from Trueblood et al.).

Usage

```
data(med_dec)
```

Format

A data frame with 11000 rows and 9 variables:

id identification number of the participant

group expertise of participant; "experienced", "inexperienced", or "novice". The first two levels refer to different type of medical professional (i.e., experts).

block block number

trial index of trial for each participant

classification true classification of the pictured cell; i.e. the correct response

difficulty adjudged difficulty of the task for the particular image

response response given by the participant; either "blast" or "non-blast"

rt the response time associated with the response, in seconds

stimulus the image file used for the specific trial

Details

At the beginning of the experiment, both novices and medical experts completed a training to familiarize themselves with blast cells. After that, each participant performed the main task in which they judged whether or not presented images were blast cells or non-blast cells. Among them, some of the cells were judged as easy and some as difficult trials by an additional group of experts. The current data set only contains the data from the "accuracy" condition (i.e., Trueblood et al. considered additional conditions that are not part of the current data set).

The relevant part of the method section for the accuracy condition from the original paper is as follows:

"The main task consisted of six blocks with 100 trials in each block. The main task was the same as the practice block, where participants were asked to identify single images. However, participants did not receive trial-by-trial feedback about their choices. They received feedback about their performance at the end of each block. The 100 trials in each block were composed of equal numbers of easy blast images, hard blast images, easy non-blast images, and hard non-blast images, fully randomized.

There were three manipulations across blocks: accuracy, speed, and bias. In the accuracy blocks, participants were instructed to respond as accurately as possible and were given 5 s to respond. [...] If they responded after the deadline, they received the message "Too Slow!" The 5-s [...] response windows for the accuracy [...] [condition was] based on the response time data from the three expert raters. The 0.975 quantile of the expert raters' response times was 4.96 s; thus, we set the accuracy response window to 5 s.

The order of the first three blocks was randomized but with the constraint that there was one block for each type of manipulation (i.e., accuracy, speed, and bias). The order of the last three blocks was identical to the order of the first three blocks."

Note that this dataset contains some negative response times that indicate a missing response (i.e., the response value for that trial is 'NA'). Take care in removing these values before using this dataset. See our Validity vignette for an example of use in an optimization setting.

Source

Trueblood, J.S., Holmes, W.R., Seegmiller, A.C. et al. The impact of speed and bias on the cognitive processes of experts and novices in medical image decision-making. *Cogn. Research* 3, 28 (2018). <https://doi.org/10.1186/s41235-018-0119-2>

Examples

```
data("med_dec", package = "fddm")
str(med_dec)

## number of participants per expertise condition:
aggregate(id ~ group, med_dec, function(x) length(unique(x)))

## number of trials per participant
aggregate(rt ~ group + id, med_dec, length)
```

Index

* **dataset**
 med_dec, 6

dfddm, 2

med_dec, 6