Bayes factor design analysis: Manual for the BFDA package

Felix Schönbrodt, Angelika Stefan

2024-11-06

This document demonstrates how to do a design analysis (aka. power analysis) for studies which use Bayes factors as index of evidence. For more details about Bayes Factor Design Analysis (BFDA), see our papers:

> Schönbrodt, F. D. & Wagenmakers, E.-J. (2017). Bayes Factor Design Analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 128-142. doi:10.3758/s13423-017-1230-y. [PDF][OSF project with reproducible code]
>
> Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E. (2018). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, 1042-1058. doi:10.3758/s13428-018-01189-8. [PDF][OSF project with reproducible code]

If you use this package to compute and report your design analysis, please cite it as:

> Schönbrodt, F. D. & Stefan, A. M. (2018). BFDA: An R package for Bayes factor design analysis (version 0.3). Retrieved from https://github.com/nicebread/BFDA

Please note that this package is still a development version; take the results with a grain of salt.

Installation

The BFDA package is not on CRAN yet, but you can install the development version from Github:

```
library(devtools)
install_github("nicebread/BFDA", subdir="package")
```

The general workflow

1. Simulate many hypothetical studies, both under H1 and under H0, using the function `BFDA.sim`
2. Analyze the simulated studies, using the function `BFDA.analyze`
3. Plot the simulated studies (`plot`, `SSD`, `evDens`)
4. Tune your design in a way that you achieve the desired goals with adequate probability (`SSD`)

To summarize, the general workflow is (here shown without parameters; these are discussed later):

```
sim.H1 <- BFDA.sim(expected.ES = 0.5, ...)
sim.H0 <- BFDA.sim(expected.ES = 0, ...)

BFDA.analyze(sim.H1)
BFDA.analyze(sim.H0)

plot(sim.H1)
plot(sim.H0)

SSD(sim.H1)
SSD(sim.H0)
```

1. Simulating hypothetical studies for a prospective design analysis

As we do not know in advance whether H1 or H0 provide a better predictive performance of real world data, we want to evaluate the performance of a design under *both* hypotheses. Hence, we have to simulate a "H1 world" and a "H0 world":

```
sim.H1 <- BFDA.sim(expected.ES=0.5, type="t.between",
                   prior=list("Cauchy",list(prior.location=0, prior.scale=sqrt(2)/2)),
                   n.min=20, n.max=300, alternative="greater", boundary=Inf, B=1000,
                   verbose=TRUE, cores=1, stepsize = 10)

sim.H0 <- BFDA.sim(expected.ES=0, type="t.between",
                   prior=list("Cauchy", list(prior.location=0, prior.scale=sqrt(2)/2)),
                   n.min=20, n.max=300, alternative="greater", boundary=Inf, B=1000,
                   verbose=TRUE, cores=1, stepsize = 10)
```

Let's go through the parameters of the simulation function:

- `expected.ES`: The simulated population effect size (design prior) which can either be a smallest effect size of interest or an expected effect size. In a classical power analysis, this would be a fixed number. Here, you can also provide a vector, which quantifies the uncertainty about the true ES. For example: `expected.ES=rnorm(100000, 0.5, 0.1)`. If a vector is provided, a new ES is drawn from this vector for each simulated study. The metric for `expected.ES` depends on the type of design (see next bullet point):
  - `type = "t.between"` or `type = "t.paired"`: expected.ES has to be provided as Cohen's d
  - `type = "correlation"`: expected.ES has to be provided as correlation
  - `type = "abtest"`: expected.ES can be either an odds ratio, log odds ratio, relative risk, or absolute risk (depending on the argument `options.sample`)
- `type`: Type of design. Currently, 4 designs are implemented: A between-group t-test (`"t.between"`), a paired t-test (`"t.paired"`), a correlation test (`"correlation"`), and an AB test (`"abtest"`)
- `prior`: This argument specifies the prior distribution under the alternative hypothesis. It consists of a list with two elements: The first element is a character vector which specifies the type of the prior distribution. The second element is a list which contains the hyperparameters of the prior distribution. For example, the prior distribution for t-tests could be defined as `prior = list("Cauchy", prior.location = 0, prior.scale = 1))`. The choice of prior distributions and the default prior settings depend on the type of design (see next bullet point)
  - `type = "t.between"` or `type = "t.paired"`: You can choose from 3 families of prior distributions on the effect size Cohen's d. The default prior is a zero-centered Cauchy distribution with a scale parameter of $\sqrt{(2)}/2$ which is equivalent to the default setting in the *BayesFactor* package.
    - A Cauchy distribution (`"Cauchy"`) with the hyperparameters `prior.location` (non-centrality parameter) and `prior.scale` (scale parameter)
    - A t distribution (`"t"`) with the hyperparameters `prior.location`, (non-centrality parameter) `prior.scale`, (scale parameter) and `prior.df`(degrees of freedom)
    - A normal distribution (`"normal"`) with the hyperparameters `prior.mean` (mean) and `prior.variance` (variance)
  - `type = "correlation"`: The prior distribution on the correlation is a stretched beta prior (`"stretchedbeta"`) with the hyperparameter κ (`prior.kappa`). The stretched beta distribution is a beta distribution with the parameters $\alpha = \beta = 1/\kappa$ whose domain is extended from [0, 1] to [-1, 1]. The default prior setting is a uniform prior (`list("stretchedbeta", list(prior.kappa = 1))`).

  - `type = "abtest"`: The prior distribution on ψ (which is equivalent to the difference in log-odds) is a normal distribution (`"normal"`) with the hyperparameters `prior.mean` (mean) and `prior.variance` (variance). The default setting is a standard normal distribution (`list("normal", list(prior.mean = 0, prior.variance = 1))`). The prior on the noise parameter β is fixed to a standard normal distribution and cannot be changed.

- `n.min` and `n.max`: The initial sample size and the maximum sample size that is tested in the sequential procedure. If the `design` argument is set to `fixed.n`, only the argument `n.max` will be considered.
- `design`: This argument specifies the experimental design which can either be a fixed-N design (`fixed.n`) or a sequential design (`sequential`). In the case of a fixed-N design, the arguments `n.min` and `boundary` are disregarded because they are irrelevant in this kind of design.
- `boundary`: The Bayes factor where a sequential run is stopped. For a fixed-N design, `boundary` is automatically set to *Inf*. You can either provide two values for a lower and an upper boundary in a vector (e.g., `c(1/3, 6)`) or only one value (e.g., `6`). If only one value is provided, the function automatically uses its reciprocal (e.g., `1/6`) for the other boundary.
- `B`: Number of simulated studies. Aim for B >= 10,000 for stable results (in this document we use B=1000 to save some computation time).
- `stepsize`: The number of observations added to the sample in each step of a sequential process. If NA, the sample is increased by 1 until a sample size of 100, for larger sample sizes it is increased by 10.
- `alternative`: Either `"two.sided"` for two-sided tests, `"greater"` for a positive directional alternative hypothesis ("the effect size is greater than zero"), or `"less"` for a negative directional alternative hypothesis ("the effect size is smaller than zero").
- `verbose`: Should the progress of the simulation be printed? TRUE/FALSE
- `cores`: Multicore support. Add as many cores as you have to speed up computations.
- `options.sample`: Further parameters passed to the data generating function in a list object. Currently only implemented for the AB test to define the type of effect size which is defined in the argument `expected.ES`. The possible parameters for the type of effect size (`list(effecttype=...)`) are `"OR"` (odds ratio), `"logOR"` (log odds ratio), `"RR"` (relative risk), and `"AR"` (absolute risk).
- `seed`: Seed that ensures reproducibility with parallel processing. If the parameter is set to `NULL`, a new seed is chosen at each run. The default seed is `1234`.

The simulations of the "H1 world" and a "H0 world" should have the same parameters except of the `expected.ES`. This means that in the actual data analysis, we will apply the same test to the data set, regardless whether data came from H0 or H1 (what we wouldn't know anyway in practice).

By default, a full sequential design without evidential stopping threshold is simulated. This means that samples are drawn in a sequential process (n.min + 1 + 1 + ...) until the maximum sample size is reached and the hypothesis test is conducted at each stage of this sequential process. The process does not stop when a Bayes factor boundary is reached (e.g., BF(accumulated sample) > 6 or BF(accumulated sample) < 1/6), but when n.max is reached. This allows to extract the results of sequential BFDA procedures with arbitrary thresholds and a maximum N of n.max. With the arguments `design` and `n.max` it is possible to change the procedure to a fixed-N BFDA (e.g., `design="fixed.n", n.max=200` for a fixed-N design with 200 observations) or an open-ended sequential procedure (e.g., `design="sequential", n.max = Inf`).

2. Analyze the simulations

Next, we can retrieve summary statistics from our simulations by applying `BFDA.analyze` to the simulation objects. If you simulated data in a fixed-N design (see argument `design` of the `BFDA.sim` function) or in a sequential design with infinite boundaries (see arguments `boundary` and `design` of the `BFDA.sim` function), you can retrieve the summary statistics for a fixed-N design.

For example, we can get the operational characteristics of a **fixed-n design** with 50 observations per group, and symmetric evidence boundaries of 1/6 and 6 by executing the code snippet below. Note that the `boundary` argument here does not define stopping boundaries for a sequential process but boundaries for "strong enough" evidence for H0 and H1, respectively.

```
BFDA.analyze(sim.H1, design="fixed", n=50, boundary=6)
```

```
## 51.8% showed evidence for H1 (BF > 6)
## 48.1% were inconclusive (0.1667 < BF < 6)
## 0.1% showed evidence for H0 (BF < 0.1667)
```

```
BFDA.analyze(sim.H0, design="fixed", n=50, boundary=6)
```

```
## 0.9% showed evidence for H1 (BF > 6)
## 63.3% were inconclusive (0.1667 < BF < 6)
## 35.8% showed evidence for H0 (BF < 0.1667)
```

The results of the analysis show in how many percent of the simulated iterations (see the `B` in the `BFDA.sim` function) the Bayes factor exceeded the upper or lower boundary, that is, in how many percent of the simualted iterations "strong enough evidence" for H0 or H1 could be achieved.

The `BFDA.analyze` function also allows you to retrieve summary statistics for a **sequential design** - in the example below for a sequential design with a minimum sample size of 20, a maximum sample size of 300, and symmetric stopping boundaries of 1/10 and 10.

```
BFDA.analyze(sim.H1, design="sequential", n.min=20, n.max=300, boundary=10)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)         0%
## 2    Studies terminating at a boundary       100%
## 3       --> Terminating at H1 boundary       100%
## 4       --> Terminating at H0 boundary         0%
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 66
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  60 100 120 150
```

Here, all studies hit a boundary before n.max is reached. The average sample size at the stopping point is 67. Additionally, the output shows quantiles of the sample size distribution. For example, one of the boundaries was hit with a sample size of 60 or less in 50% of the time.

If we reduce n.max, some studies do not reach an evidential threshold. In this case, the output shows additional information about the studies ending at the maximum sample size (n.max = 100).

```
BFDA.analyze(sim.H1, design="sequential", n.min=20, n.max=100, boundary=10)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=100)        16%
## 2    Studies terminating at a boundary        84%
## 3       --> Terminating at H1 boundary        84%
## 4       --> Terminating at H0 boundary         0%
##
## Of 16% of studies terminating at n.max (n=100):
## 7.9% showed evidence for H1 (BF > 3)
## 7.9% were inconclusive (3 > BF > 1/3)
## 0.2% showed evidence for H0 (BF < 1/3)
##
```

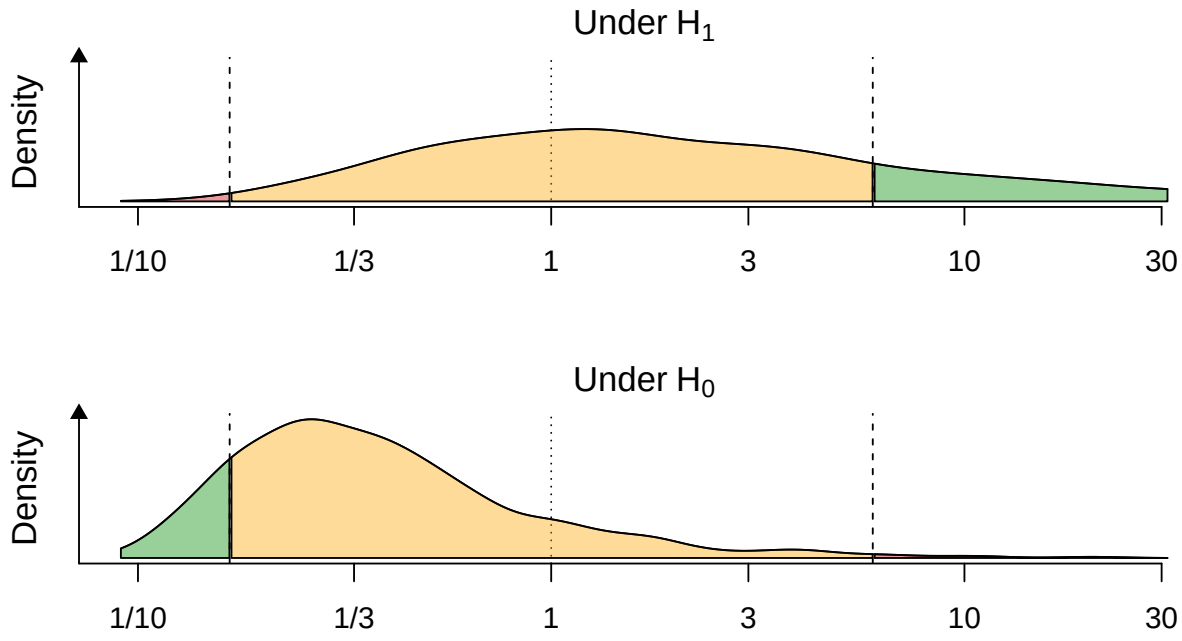
```
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 59
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  60 100 100 100
```

3. Plot the design analysis

Compare distributions of BFs for a fixed n

The following function gives a graphic representation of the distribution of Bayes factors for a fixed-N design. Bayes factors showing "strong enough" evidence (as defined by the `boundary` argument) in the right direction are shown in green, Bayes factors showing "strong enough" evidence in the wrong direction are shown in red. Inconclusive evidence is indicated in yellow color.

```
evDens(BFDA.H1=sim.H1, BFDA.H0=sim.H0, n=20, boundary=c(1/6, 6), xlim=c(1/11, 31))
```

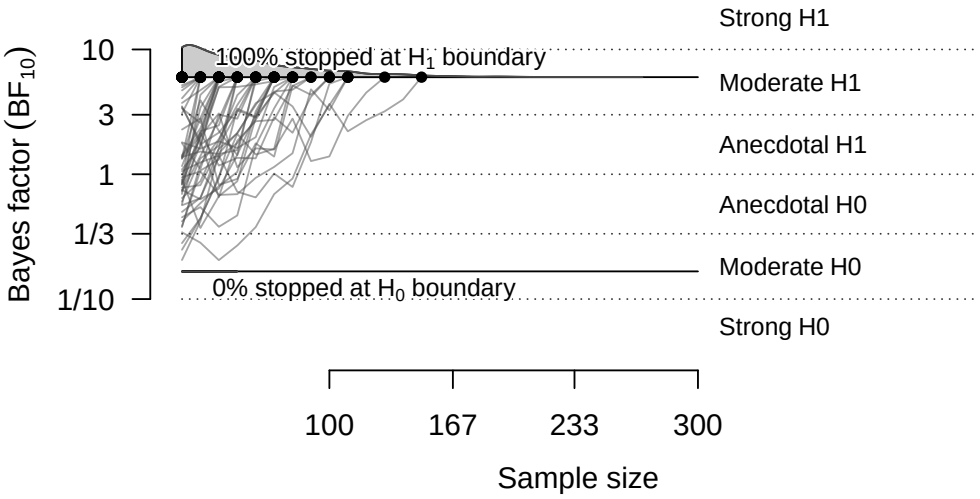


Open-ended sequential design

The following function gives a graphic summary of an open-ended sequential process, that is a sequential process without a maximum sample size. For didactic purposes, we will apply it to our simulated dataset which was simulated with a maximum sample size of N = 300 (since all sequential processes reached a boundary before n.max was reached, we will pretend the data set was simulated with n.max = Inf).

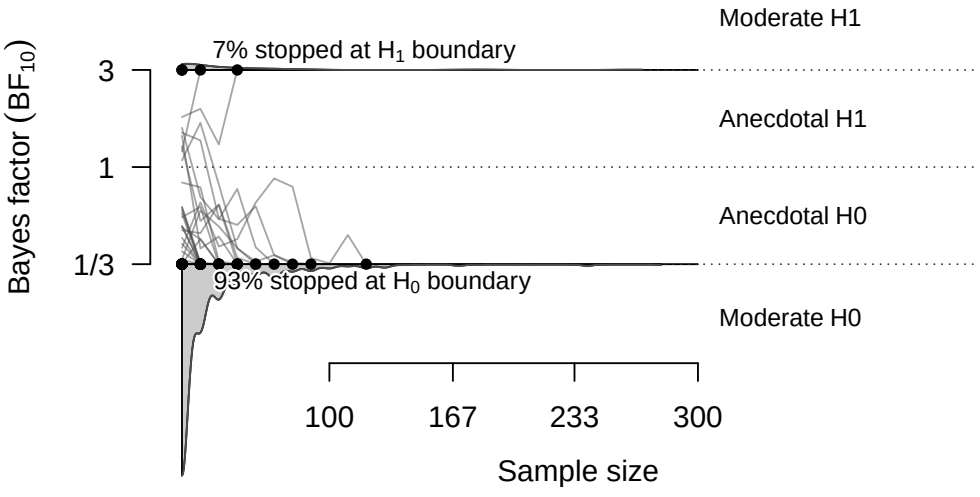
Under H1:

```
plot(sim.H1, n.min=20, boundary=c(1/6, 6), n.trajectories = 60)
```



Under H0:

```
plot(sim.H0, n.min=20, boundary=c(1/3, 3))
```



In the middle, the figures show several sampling trajectories. The number of these trajectories can be controlled with the argument `n.trajectories`. By default, the selected trajectories reflect the proportions of each

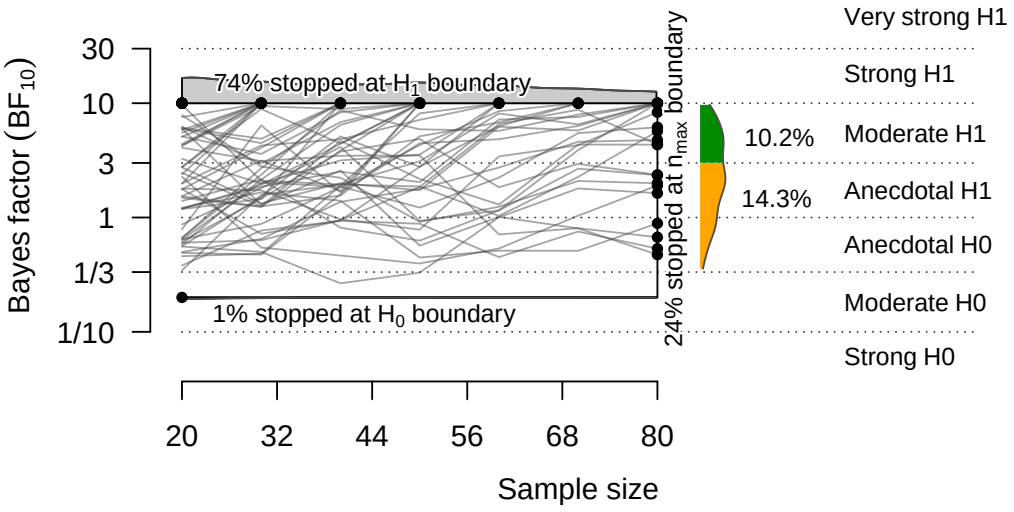
stopping category (upper/lower/n.max hits). Alternatively, a fixed set of trajectories can be shown (argument `traj.selection = "fixed"`). Below and above the boundaries, density curves show the distribution of sample sizes at which the trajectories arrive at the respective boundary and the overall percentage of trajectories arriving at the respective boundary is indicated. The horizontal lines show evidential categories as proposed by Jeffreys (1961).

Sequential design with n.max and asymmetric boundaries

The plot function can also be used to illustrate a sequential BFDA with a maximum sample size. To do so, specify the argument `n.max`. Here, we additionally use asymmetric stopping boundaries (see the `boundary` argument).

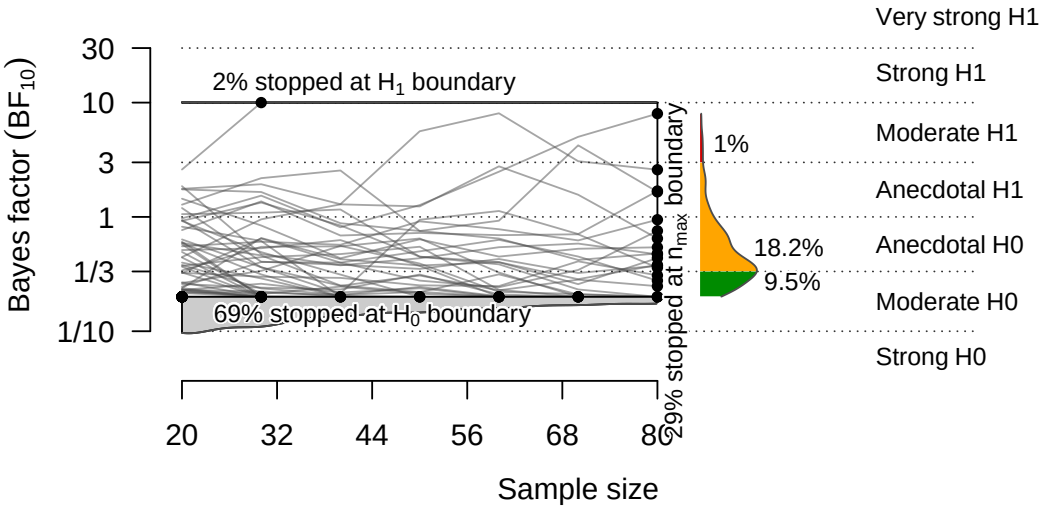
Under H1:

```
plot(sim.H1, n.min=20, n.max=80, boundary=c(1/5, 10))
```



Under H0:

```
plot(sim.H0, n.min=20, n.max=80, boundary=c(1/5, 10), forH1 = FALSE)
```



In the plots for sequential designs with a maximum sample size a distribution is added on the right side of the plots that illustrates the distribution of Bayes factors in sampling trajectories that arrived at n.max. For example, we can see in the upper plot that roughly 10% of the trajectories arriving at n.max had a Bayes factor of larger than 3 in favor of the alternative hypothesis in the end.

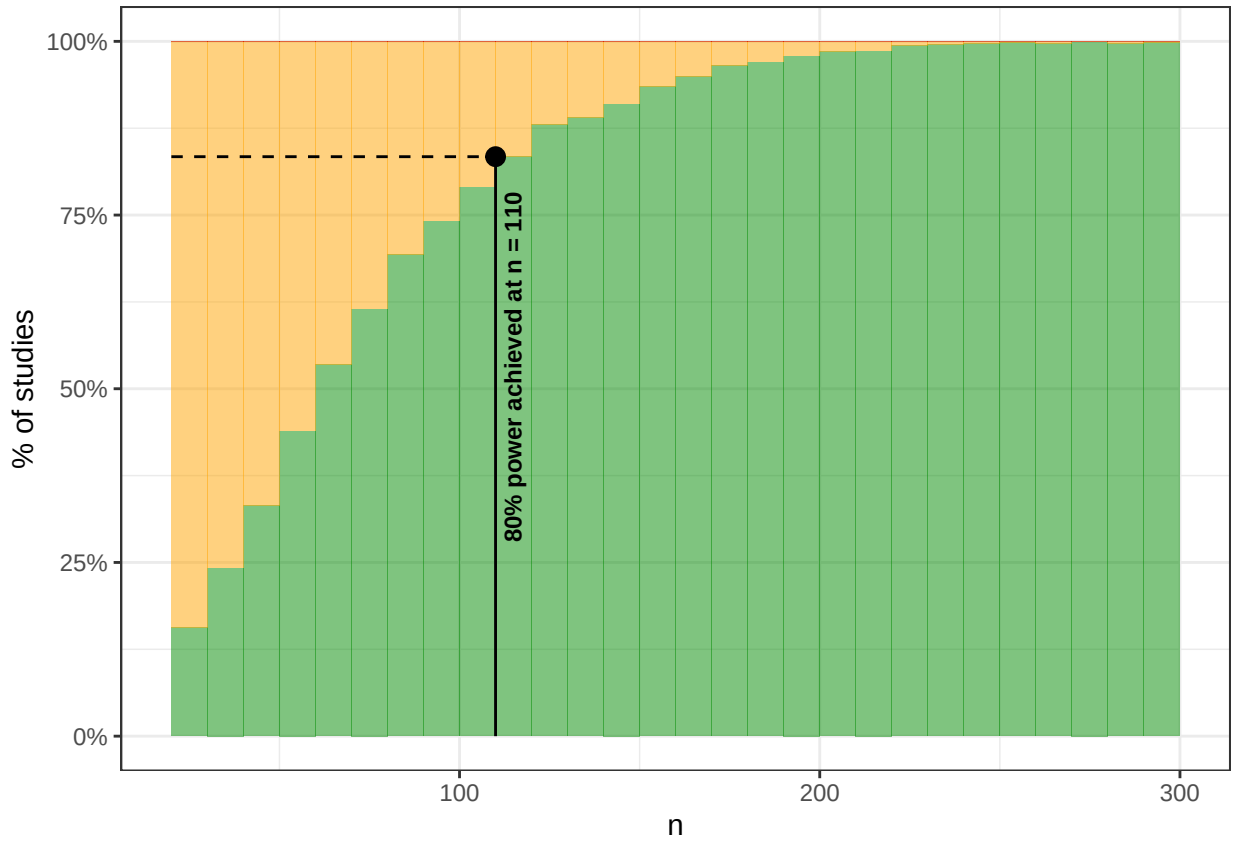
4. Sample Size Determination (SSD)

The SSD function helps you to find the right sample size in a fixed-N design.

What sample size do you need to ensure, say, 80% probability that a study design finds an effect of size of 0.5, that is, that at least 80% of resulting Bayes factors are equal or larger than the upper boundary (here: BF >= 10) when the population effect size is 0.5?

```
SSD(sim.H1, power=.80, boundary=c(1/10, 10))
```

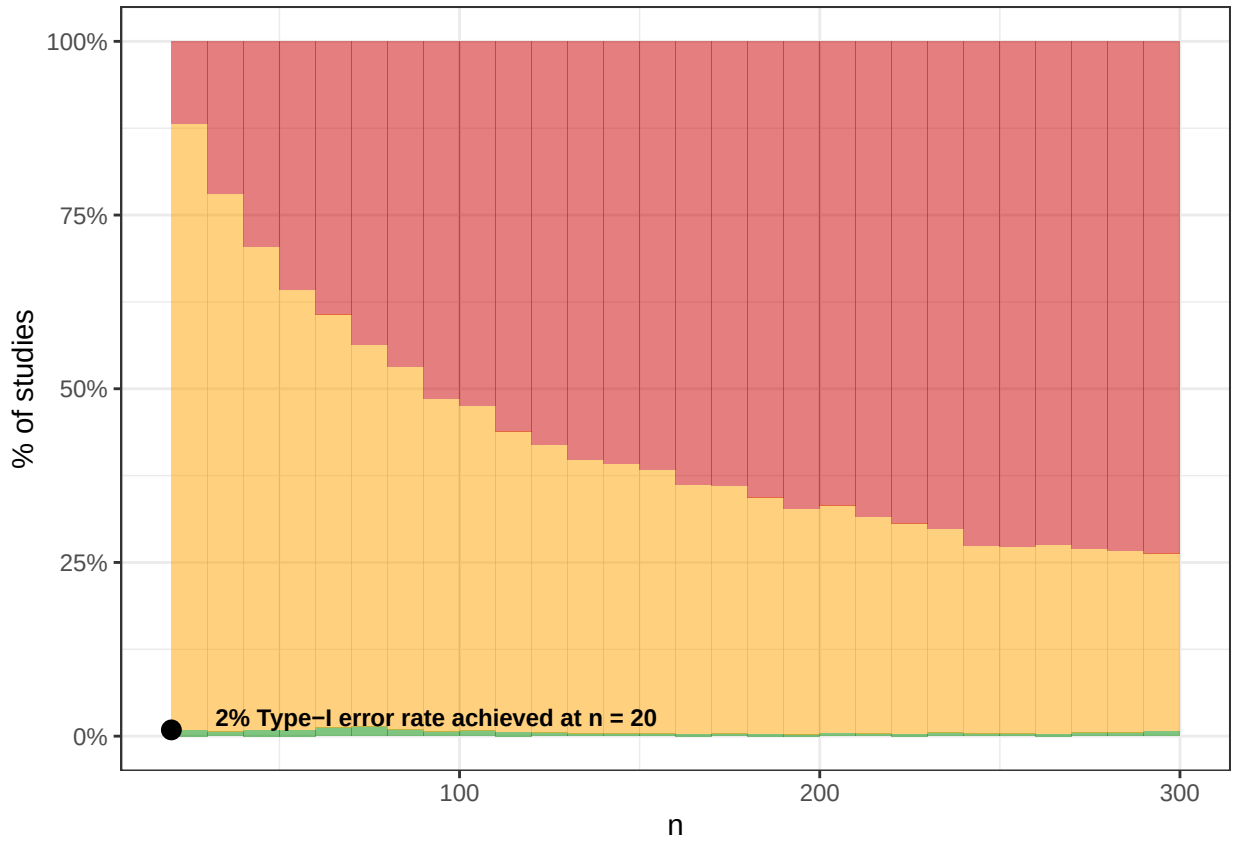
```
## Sample size determination for a fixed-n design:
## ---------------------------------------
##
## A >= 80% (actual: 83.4%) power achieved at n = 110
## This setting implies long-term rates of:
## 16.6% inconclusive results and
##    0% false-negative results.
```



What sample size do I need to have less than 2% of studies with a false positive error, i.e., Bayes factors equal or larger than the upper boundary (here: BF >= 6) when H0 is true in the population? Note that you should never use this as the only criterion to select a fixed sample size because it does not ensure strong evidence (e.g., with a sample size of 20, less than 2% of studies yield a false positive result, but also only 14.9% of the Bayes factors are smaller or equal 1/6).

```
SSD(sim.H0, alpha=.02, boundary=c(1/6, 6))
```

```
## Sample size determination for a fixed-n design:
## ---------------------------------------
##
## A >= 2% (actual: 0.9%) long-term rate of Type-I errors is achieved at n = 20
## This setting implies long-term rates of:
## 87.2% inconclusive results and
##    11.9% true-negative results.
```



Note: The SSD function automatically detects whether a H1 or a H0 simulation is analyzed. Also note that these numbers of necessary sample sizes are based on simulations and depend on the seed used in the simulation function. The larger the number `B` of simulations, the less variable are these estimates.

Paired t-test: A complete example

The BFDA uses informed t-test functions by Gronau, Ly, & Wagenmakers (2019) to compute the between and paired t-test. By default, the prior on effect size under H1 is defined as a central Cauchy distribution with a scale parameter of sqrt(2)/2 as it is used in the *BayesFactor* package. The effect size metric is Cohen's d (standardized difference of means).

```
#devtools::install_github("nicebread/BFDA", subdir="package")
#library(BFDA)

# do a sequential design analysis
s1 <- BFDA.sim(expected.ES=0.4,
               prior=list("t", list(prior.location=0, prior.scale=sqrt(2)/2, prior.df=1)),
               n.min=50, stepsize=5, n.max=300, type="t.paired", design="sequential",
               alternative="greater", B=1000, cores=1, verbose=FALSE)
s0 <- BFDA.sim(expected.ES=0,
               prior=list("t", list(prior.location=0, prior.scale=sqrt(2)/2, prior.df=1)),
               n.min=50, stepsize=5, n.max=300, type="t.paired", design="sequential",
               alternative="greater", B=1000, cores=1, verbose=FALSE)

# if no n.min and n.max is provided in the `BFDA.analyze` function,
# the values from the simulation are taken
```

```r
BFDA.analyze(s1, design="sequential", boundary=10)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)         0%
## 2    Studies terminating at a boundary       100%
## 3       --> Terminating at H1 boundary       100%
## 4       --> Terminating at H0 boundary         0%
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 66
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  50  80 100 120
```

```r
BFDA.analyze(s0, design="sequential", boundary=10)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)      13.2%
## 2    Studies terminating at a boundary      86.8%
## 3       --> Terminating at H1 boundary       1.4%
## 4       --> Terminating at H0 boundary      85.4%
##
## Of 13.2% of studies terminating at n.max (n=300):
## 0.2% showed evidence for H1 (BF > 3)
## 5.4% were inconclusive (3 > BF > 1/3)
## 7.6% showed evidence for H0 (BF < 1/3)
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 127
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  85 220 300 300
```

```r
BFDA.analyze(s1, design="sequential", boundary=6)
```

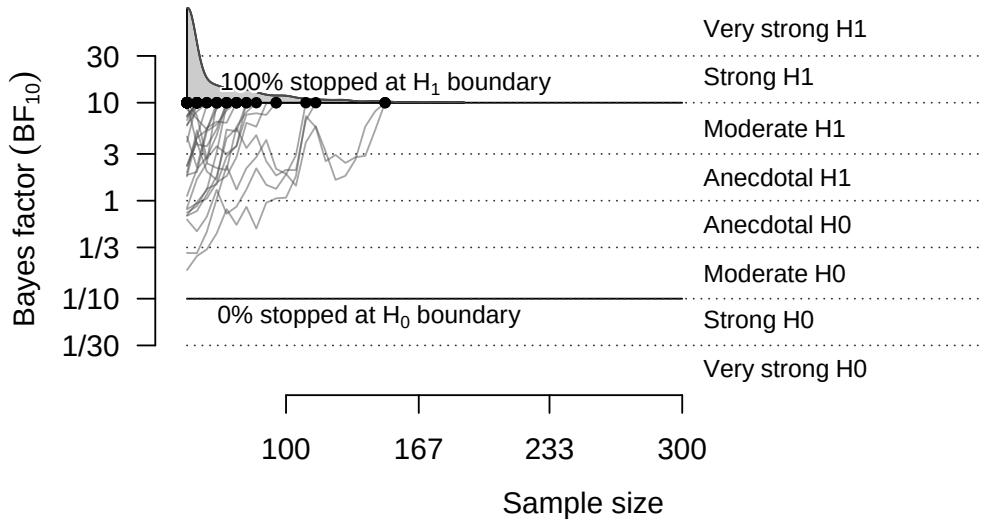
```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)         0%
## 2    Studies terminating at a boundary       100%
## 3       --> Terminating at H1 boundary      99.1%
## 4       --> Terminating at H0 boundary       0.9%
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 61
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  50  70  86 105
```

```r
BFDA.analyze(s0, design="sequential", boundary=6)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)       3.6%
## 2    Studies terminating at a boundary      96.4%
## 3       --> Terminating at H1 boundary       2.3%
## 4       --> Terminating at H0 boundary      94.1%
##
## Of 3.6% of studies terminating at n.max (n=300):
```

```
## 0% showed evidence for H1 (BF > 3)
## 2.8% were inconclusive (3 > BF > 1/3)
## 0.8% showed evidence for H0 (BF < 1/3)
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 82
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  50 100 165 250
```

```
plot(s1)
```

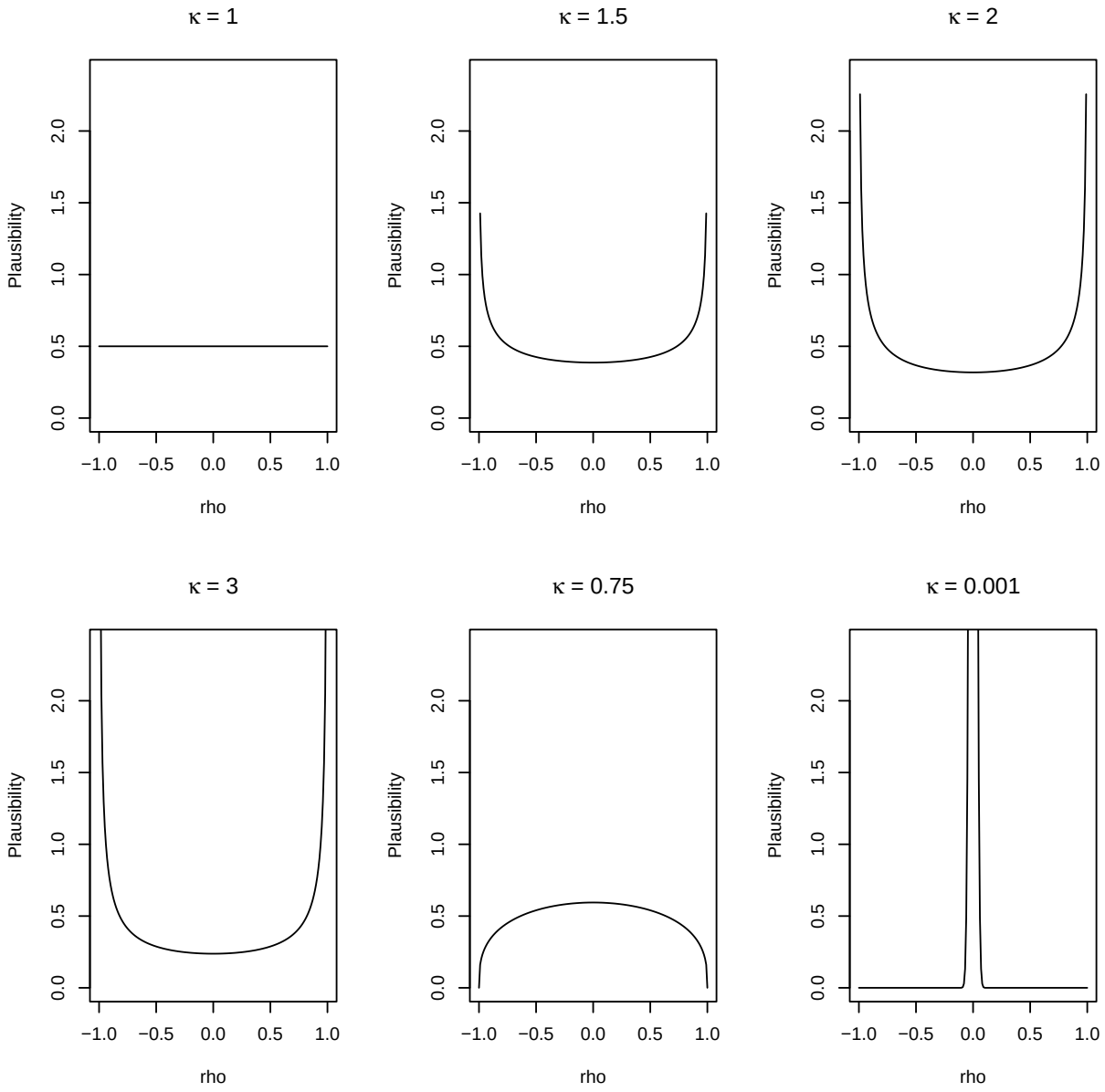


Correlation: A complete example

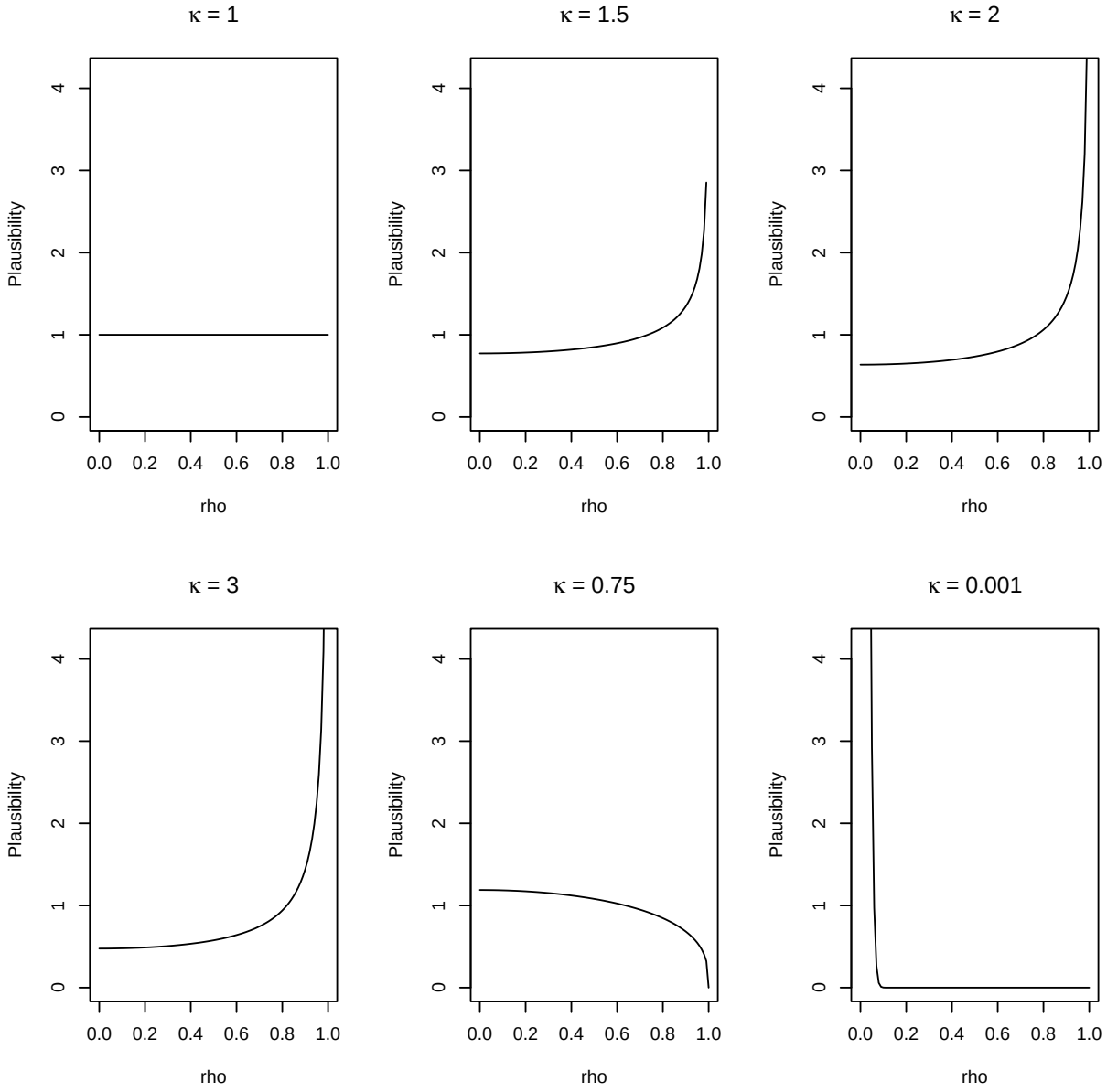
The BFDA uses the source code of the statistics software JASP (available on Github) which is based on a paper by Ly, Verhagen & Wagenmakers (2016) and a paper by Ly, Marsman, & Wagenmakers. The kappa parameter which can be passed to the BFDA.sim function as part of the "prior" argument, corresponds to the "beta prior width" parameter in JASP. By default, the BFDA uses a κ parameter of 1, which is equivalent to a uniform prior distribution over the possible correlations.

A short explanation of the stretched beta prior: First, the correlation is rescaled to lie between 0 and 1, then a beta distribution with the parameters $\alpha = \beta = 1/\kappa$ is assigned to it. Then, the beta distribution is transformed to the (-1, 1) scale and the Bayes factors are calculated as a function of κ. When $\kappa = 1$, this corresponds to a uniform prior on the correlation coefficient, as in Jeffreys's default analysis. When $\kappa \to 0$, H1 becomes indistinguishable from H0 and consequently the Bayes factor is 1. Values of κ in between 0 and ∞ define an continuous range of different alternative hypotheses that represent different beliefs about the extent to which large values for the correlation are plausible.

Here are some plots for different settings of kappa for the two-sided case:



Here are some plots for different settings of kappa for the one-sided case with a positive directional hypothesis:



Here is a complete walkthrough for a one-sided (positive directional) correlation BF design analysis, which places prior weight on smaller effect sizes (kappa = 2):

```
#devtools::install_github("nicebread/BFDA", subdir="package")
#library(BFDA)

# do a sequential design analysis
c1 <- BFDA.sim(expected.ES=0.21, prior=list("stretchedbeta",list(prior.kappa=2)),
               n.min=50, stepsize=10, n.max=300, B=1000, type="correlation",
               design="sequential", alternative="greater", cores=1, verbose=FALSE)
c0 <- BFDA.sim(expected.ES=0, prior=list("stretchedbeta", list(prior.kappa=2)),
               n.min=50, stepsize=10, n.max=300, B=1000, type="correlation",
               design="sequential", alternative="greater", cores=1, verbose=FALSE)
```

```r
# if no n.min and n.max is provided in the `BFDA.analyze` function,
# the values from the simulation are taken
BFDA.analyze(c1, design="sequential", boundary=10)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)      12.5%
## 2    Studies terminating at a boundary      87.5%
## 3       --> Terminating at H1 boundary      76.9%
## 4       --> Terminating at H0 boundary      10.6%
##
## Of 12.5% of studies terminating at n.max (n=300):
## 4.6% showed evidence for H1 (BF > 3)
## 7.7% were inconclusive (3 > BF > 1/3)
## 0.2% showed evidence for H0 (BF < 1/3)
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 157
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
## 140 250 300 300
```

```r
BFDA.analyze(c0, design="sequential", boundary=10)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)       7.4%
## 2    Studies terminating at a boundary      92.6%
## 3       --> Terminating at H1 boundary       0.8%
## 4       --> Terminating at H0 boundary      91.8%
##
## Of 7.4% of studies terminating at n.max (n=300):
## 0.2% showed evidence for H1 (BF > 3)
## 2.5% were inconclusive (3 > BF > 1/3)
## 4.7% showed evidence for H0 (BF < 1/3)
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 103
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  60 140 250 300
```

```r
BFDA.analyze(c1, design="sequential", boundary=6)
```

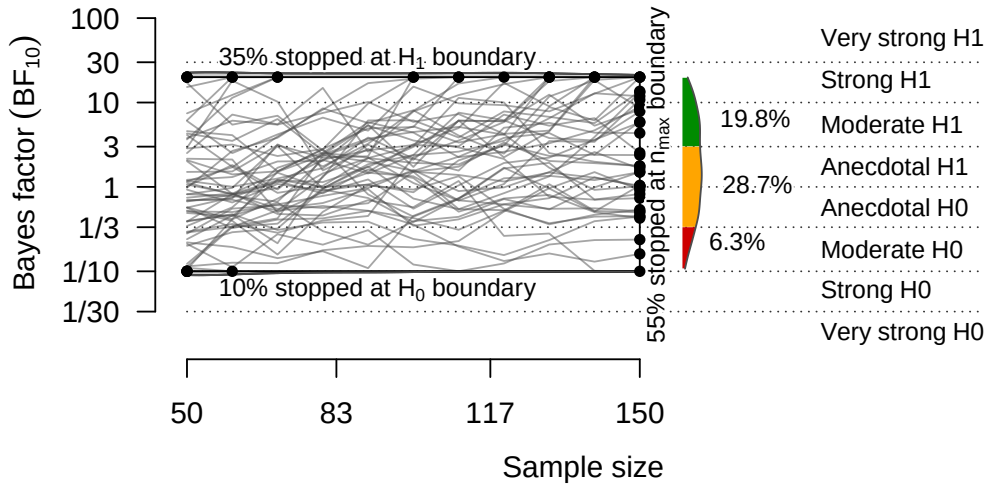
```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)       4.4%
## 2    Studies terminating at a boundary      95.6%
## 3       --> Terminating at H1 boundary      70.6%
## 4       --> Terminating at H0 boundary        25%
##
## Of 4.4% of studies terminating at n.max (n=300):
## 1% showed evidence for H1 (BF > 3)
## 3.4% were inconclusive (3 > BF > 1/3)
## 0% showed evidence for H0 (BF < 1/3)
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 115
##
```

```
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  90 180 230 280
```

```
BFDA.analyze(c0, design="sequential", boundary=6)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)         2%
## 2    Studies terminating at a boundary        98%
## 3       --> Terminating at H1 boundary       1.3%
## 4       --> Terminating at H0 boundary      96.7%
##
## Of 2% of studies terminating at n.max (n=300):
## 0% showed evidence for H1 (BF > 3)
## 0.9% were inconclusive (3 > BF > 1/3)
## 1.1% showed evidence for H0 (BF < 1/3)
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 71
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  50  80 120 181
```

```
plot(c1, boundary=c(1/10, 20), n.max=150)
```



AB test: A complete example

The Bayesian AB test in the BFDA package is based on the *abtest* R-package (Gronau, Raj & Wagenmakers, 2019). The test uses a normal prior on the difference of log odds as an analysis prior. The default is a standard normal distribution (mean = 0, variance = 1). In the BFDA, you can use an odds ratio, a log odds ratio, a relative risk, or an absolute risk to define a design prior on the population effect size (arguments

`expected.ES` and `options.sample`).

$Odds\ Ratio = \frac{p2}{1-p2} / \frac{p1}{1-p1}$

$log\,Odds\ Ratio = log(Odds\ Ratio)$

$Relative\ Risk = \frac{p2}{p1}$

$Absolute\ Risk = p2 - p1$

```
#devtools::install_github("nicebread/BFDA", subdir="package")
#library(BFDA)

# do a sequential design analysis
ab1 <- BFDA.sim(expected.ES=2, prior=list("normal", list(prior.mean=0.5, prior.variance = 1)),
                n.min=50, stepsize=10, n.max=300, B=1000, type="abtest",
                design="sequential", alternative="two.sided", cores=1, verbose=FALSE,
                options.sample = list(effecttype = "OR"))
ab0 <- BFDA.sim(expected.ES=0, prior=list("normal", list(prior.mean=0.5, prior.variance = 1)),
                n.min=50, stepsize=10, n.max=300, B=1000, type="abtest",
                design="sequential", alternative="two.sided", cores=1, verbose=FALSE,
                options.sample = list(effecttype = "OR"))

# if no n.min and n.max is provided in the `BFDA.analyze` function,
# the values from the simulation are taken
BFDA.analyze(ab1, design="sequential", boundary=10)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)       5.9%
## 2    Studies terminating at a boundary      94.1%
## 3       --> Terminating at H1 boundary      94.1%
## 4       --> Terminating at H0 boundary         0%
##
## Of 5.9% of studies terminating at n.max (n=300):
## 2.5% showed evidence for H1 (BF > 3)
## 3.1% were inconclusive (3 > BF > 1/3)
## 0.3% showed evidence for H0 (BF < 1/3)
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 134
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
## 110 210 260 300
```

```
BFDA.analyze(ab0, design="sequential", boundary=10)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)         0%
## 2    Studies terminating at a boundary       100%
## 3       --> Terminating at H1 boundary       100%
## 4       --> Terminating at H0 boundary         0%
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 50
##
```

```
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  50  50  50  50
```

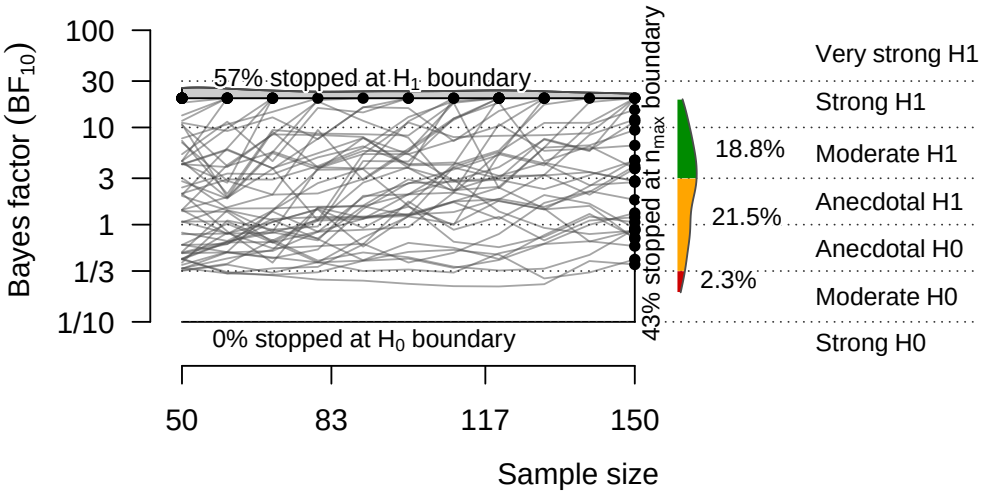
```
BFDA.analyze(ab1, design="sequential", boundary=6)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)       3.6%
## 2    Studies terminating at a boundary      96.4%
## 3       --> Terminating at H1 boundary      96.2%
## 4       --> Terminating at H0 boundary       0.2%
##
## Of 3.6% of studies terminating at n.max (n=300):
## 0.7% showed evidence for H1 (BF > 3)
## 2.8% were inconclusive (3 > BF > 1/3)
## 0.1% showed evidence for H0 (BF < 1/3)
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 119
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
## 100 180 240 290
```

```
BFDA.analyze(ab0, design="sequential", boundary=6)
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)         0%
## 2    Studies terminating at a boundary       100%
## 3       --> Terminating at H1 boundary       100%
## 4       --> Terminating at H0 boundary         0%
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 50
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  50  50  50  50
```

```
plot(ab1, boundary=c(1/10, 20), n.max=150)
```



Use case: Apply for a grant with sequential sampling

Granting agencies usually expect a fixed sample size in the planning stage in order to quantify the amount of funding. For how much participant renumeration should one apply when using a sequential design?

We suggest to determine the requested sample size using two different design analyses:

1. Compute an open-ended SBF design with the expected effect size to get a distribution of sample sizes at stopping point.
2. Compute the 80% quantile of stopping-ns: `n_q80`
3. Evaluate the characteristics of a *SBF+maxN* design with `n.max = n_q80`. Does it have acceptable false positive and false negative error rates? (If not: tune your boundaries). What is the mean and median expected sample size?
4. Apply for a sample size of `n_q80`.

```
# We use the simulation from above.
# Check the expected sample sizes for an evidential boundary of 10
a1 <- BFDA.analyze(sim.H1, design="sequential", n.min=20, boundary=10)

# --> see 80% quantile in output
a1
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=300)         0%
## 2    Studies terminating at a boundary       100%
## 3       --> Terminating at H1 boundary       100%
## 4       --> Terminating at H0 boundary         0%
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 66
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
```

```
##  60 100 120 150
```

```
# Alternative approach: access stopping-ns directly
n_q80 <- ceiling(quantile(a1$endpoint.n, prob=.80))
n_q80
```

```
## 80%
## 100
```

80% of all studies stop earlier than n = 100. How does a design with that n.max perform concerning rates of misleading evidence?

```
a2.H1 <- BFDA.analyze(sim.H1, design="sequential", n.min=20, n.max=n_q80, boundary=10)
a2.H0 <- BFDA.analyze(sim.H0, design="sequential", n.min=20, n.max=n_q80, boundary=10)
a2.H1
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=100)        16%
## 2    Studies terminating at a boundary        84%
## 3       --> Terminating at H1 boundary        84%
## 4       --> Terminating at H0 boundary         0%
##
## Of 16% of studies terminating at n.max (n=100):
## 7.9% showed evidence for H1 (BF > 3)
## 7.9% were inconclusive (3 > BF > 1/3)
## 0.2% showed evidence for H0 (BF < 1/3)
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 59
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
##  60 100 100 100
```

```
a2.H0
```

```
##                                outcome percentage
## 1 Studies terminating at n.max (n=100)      65.8%
## 2    Studies terminating at a boundary      34.2%
## 3       --> Terminating at H1 boundary         2%
## 4       --> Terminating at H0 boundary      32.2%
##
## Of 65.8% of studies terminating at n.max (n=100):
## 0.2% showed evidence for H1 (BF > 3)
## 18.5% were inconclusive (3 > BF > 1/3)
## 47.1% showed evidence for H0 (BF < 1/3)
##
## Average sample number (ASN) at stopping point (both boundary hits and n.max): n = 89
##
## Sample number quantiles (50/80/90/95%) at stopping point:
## 50% 80% 90% 95%
## 100 100 100 100
```

In this design analysis for the *SBF+maxN* design, we can see that, although we apply for 100 participants in each group, we can expect to stop with 59 participants or less with a 50% chance, if H1 is true. The false negative rate is virtually 0%. That means, if the effect exists in the expected size, this design virtually guarantees to detect it, and we have a good chance to be more efficient.

Under H0, at least half of the studies will have to use the full requested sample of 100 participants. On

average, samples will have a size of n=89 under H0. We have a 2% false positive error rate, and 32.2% of all studies will correctly stop at the H0 boundary. The remaining 65.8% of all studies will remain inconclusive with respect to the desired evidential threshold of $BF_{10} <= 1/10$. However, the Bayes factor of these studies can still be interpreted in size and direction. From the output, you can see that the majority of these inconclusive BFs still points into the correct (H0) direction.

Reproducible code for the vignette The reproducible code for the vignette can be found on [https://github.com/nicebread/BFDA].