Data auditing with jfa: : **CHEAT SHEET**

Basics

jfa is an R package that facilitates data auditing.

The package provides two functions that allow users to easily apply Bayesian or classical probability theory in their data audit workflow.

Installation

Installing the package can be done via:
`install.packages('jfa')`

Loading the package can be done via:
`library(jfa)`

Example

The blue code blocks next to the function descriptions provide a working example of the intended use.

The data for this example can be loaded via:
`data('sinoForest')`

Test the distribution of leading or last digits against Benford's law

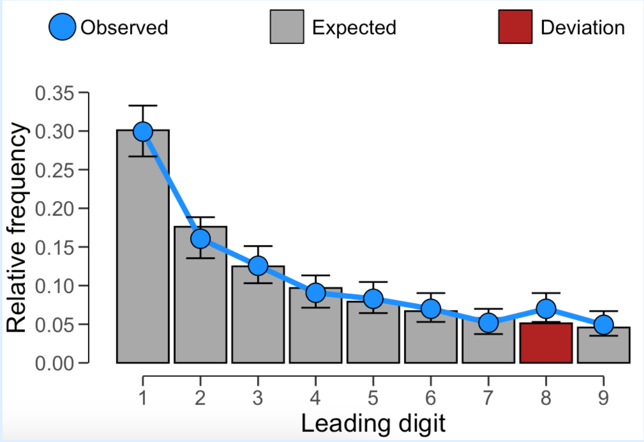
`jfa::digit_test()`

This function extracts and performs a test of the distribution of leading (two) or last digits in a vector against a reference distribution (e.g., Benford's law, the uniform distribution or a custom distribution). The `prior` argument can be used to specify a prior distribution to perform Bayesian inference.

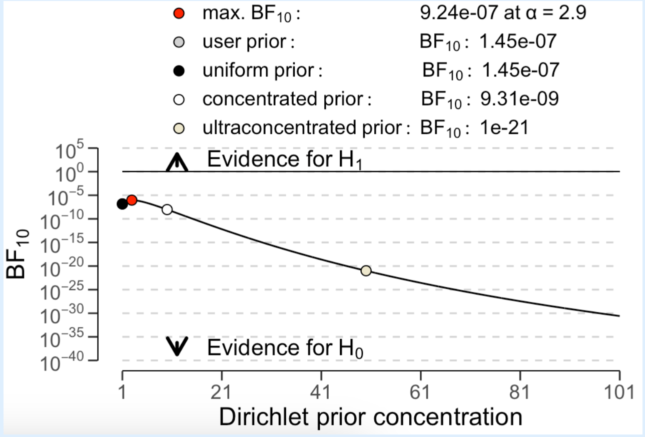
- `check:` Specifies the digits to be checked against the `reference` distribution.
- `reference`: Specifies the reference distribution to test the digits against.
- `conf.level:` Specifies the confidence level used in the analysis.
- `prior:` Specifies the concentration parameter of the Dirichlet prior distribution.

```
digit_test(sinoForest$value,
           check = 'first',
           reference = 'benford',
           conf.level = 0.95,
           prior = FALSE)
```

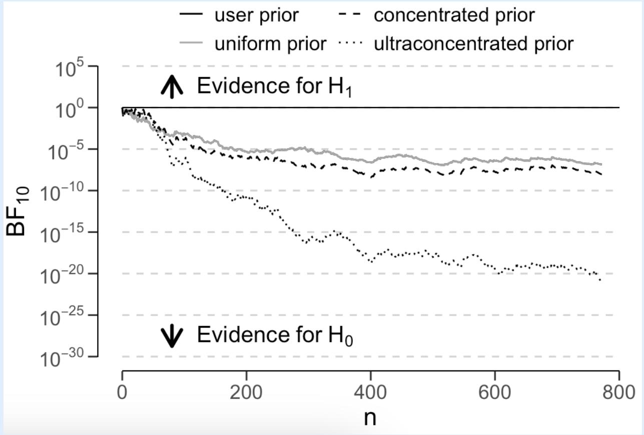
`plot(..., type = 'estimates')`



`plot(..., type = 'robustness')`



`plot(..., type = 'sequential')`



Test the frequency of repeated values for unusually high amounts

`jfa::repeated_test()`

This function analyzes the frequency with which values get repeated within a set of numbers. Unlike Benford's law, and its generalizations, this approach examines the entire number at once, not only the first or last digit(s).

- `check:` Specifies the digits to be shuffled during the analysis.
- `method`: Specifies which statistic is used to quantify the repeated values. Defaults to `af` for average frequency, but can also be `entropy` for entropy.
- `samples:` Specifies the number of samples used to bootstrap the *p*-value.

```
repeated_test(sinoForest$value,
              check = 'last',
              method = 'af',
              samples = 2000)
```

RStudio® is a trademark of RStudio, Inc. • CC BY SA Koen Derks • K.Derks@nyenrode.nl • https://koenderks.github.io/jfa • Learn more from the **webpage** or **vignettes** • Package version 0.7.2 • Updated: 2023-11