bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework

09 June 2019

Introduction

The Bayesian framework for statistics is quickly gaining in popularity among scientists, for reasons such as reliability and accuracy (particularly in noisy data and small samples), the possibility of incorporating prior knowledge into the analysis and the intuitive interpretation of results (Andrews & Baguley, 2013; Etz & Vandekerckhove, 2016; Kruschke, 2010; Kruschke, Aguinis, & Joo, 2012; Wagenmakers et al., 2018). Adopting the Bayesian framework is more of a shift in the paradigm than a change in the methodology; All the common statistical procedures (t-tests, correlations, ANOVAs, regressions, etc.) can also be achieved within the Bayesian framework. One of the core difference is that in the *frequentist* view, the effects are fixed (but unknown) and data are random. On the other hand, instead of having single estimates of the "true effect", the Bayesian inference process computes the probability of different effects *given the observed data*, resulting in a distribution of possible values for the parameters, called the *posterior distribution*.

Effects in the Bayesian framework can be described by characterizing their posterior distribution. Commonly reported indices include measures of centrality (e.g., the median, mean or MAP estimate) and uncertainty (the *credible* interval - CI). Cum grano salis, these are considered the counterparts to the coefficient point-estimates and confidence intervals of the frequentist framework. Additionally, **bayestestR** also focuses on implementing a Bayesian null-hypothesis testing framework (in a Bayesian sense, i.e., extended to general testing of "effect existence") by providing access to both established and exploratory indices of effect *existence* and *significance* (such as the Bayes factor, Morey & Rouder, 2011; the ROPE, Kruschke & Liddell, 2018; the MAP-based p-value, Mills, 2018, or the Probability of Direction - *pd*).

Existing R packages allow users to easily fit a large variety of models and extract and visualize the posterior draws. However, most of these packages only return a limited set of indices (e.g., point-estimates and CIs). **bayestestR** provides a comprehensive and consistent set of functions to analyze and describe posterior distributions generated by a variety of models objects, including popular modeling packages such as **rstanarm** (Goodrich, Gabry, Ali, & Brilleman, 2018), **brms** (Bürkner & others, 2017) or **BayesFactor** (Morey & Rouder, 2018), thus appearing as a useful tool supporting the usage and developpment of Bayesian statistics. The main functions are described below, and a full documentation is available on the package's website.

Examples of Features

The following demonstration of functions is accompanied by figures to illustrate the conceptional ideas behind the related indices. However, **bayestestR** functions also include plotting capabilities via the **see** package (Lüdecke et al., 2019a).

Indices of Centrality: Point-estimates

bayestestR offers two functions to compute point-estimates from posterior distributions: `map_estimate()` and `point_estimate()`, the latter providing options to calculate the mean, median or MAP estimate of a posterior distribution. `map_estimate()` is a convenient function to calculate the **Maximum A Posteriori** (MAP) estimate directly.

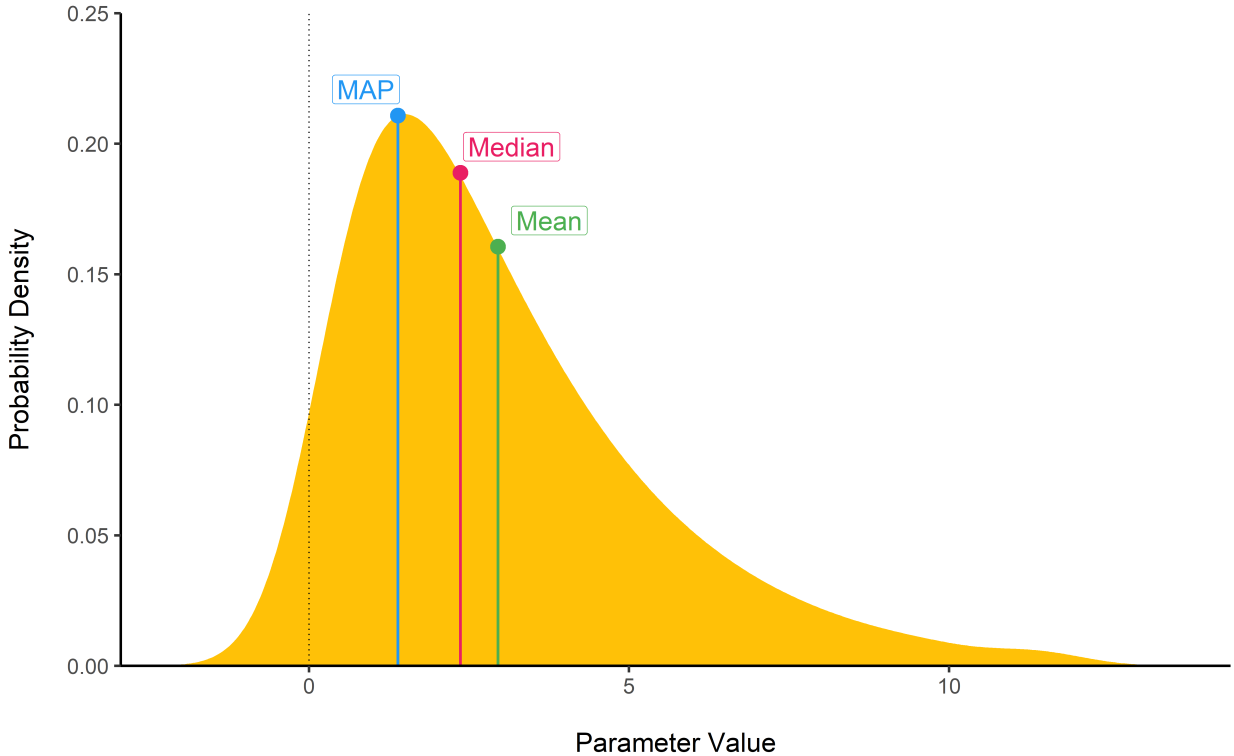

Figure 1: Indices of centrality of the posterior distribution that are often used as point-estimates: the mean (in green), median (in red), and MAP estimate (in blue)

The **posterior mean** minimizes expected *squared* error, whereas the **posterior median** minimizes expected *absolute* error (i.e. the difference of estimates from true values over samples). The **MAP estimate** is the most probable value of a posterior distribution.

```r
set.seed(1)
posterior <- rchisq(100, 3)
map_estimate(posterior)
#> MAP = 1.46

point_estimate(posterior)
#> Median = 2.31

point_estimate(posterior, centrality = "mean")
#> Mean = 2.96

point_estimate(posterior, centrality = "map")
#> MAP = 1.46
```

Quantifying Uncertainty: The Credible Interval (CI)

In order to measure the uncertainty associated with the estimation, **bayestestR** provides two functions: `eti()`, the Equal-Tailed Interval **ETI**, and `hdi()`, the **Highest Density Interval (HDI)**. Both indices (accessible via the `method` argument in the `ci()` function) can be used in the context of Bayesian posterior characterisation as **Credible Interval (CI)**.

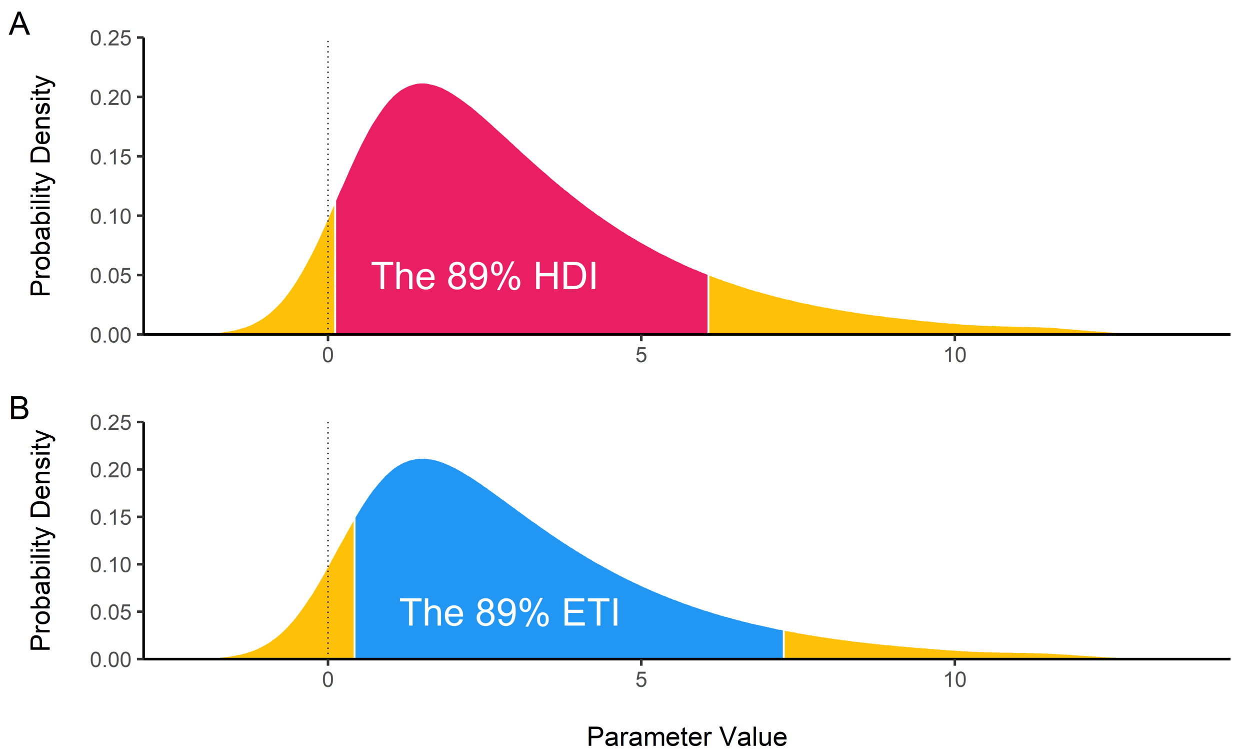

Figure 2: Indices of uncertainty that can be used as Credible Intervals (CIs): (A) the 89% Highest Density Interval (HDI); (B) the 89% Equal-Tailed Interval (ETI).

`hdi()` computes the HDI of a posterior distribution, i.e., the interval which contains all points within the interval have a higher probability density than points outside the interval. HDIs have a particular property: Unlike an equal-tailed interval (computed by `eti()`) that typically exclude 2.5% from each tail of the distribution, the HDI is *not* equal-tailed and therefore always includes the mode(s) of posterior distributions.

By default, `hdi()` and `eti()` return the 89% intervals (`ci = 0.89`), deemed to be more stable than, for instance, 95% intervals. An effective sample size of at least 10.000 is recommended if 95% intervals should be computed (Kruschke, 2015). Moreover, 89 is the highest prime number that does not exceed the already unstable and arbitrary 95% threshold (McElreath, 2018).

```
hdi(posterior)
#> # Highest Density Interval
#>
#>       89% HDI
#>  [0.11, 6.05]

eti(posterior)
#> # Equal-Tailed Interval
#>
#>       89% ETI
#>  [0.42, 7.27]
```

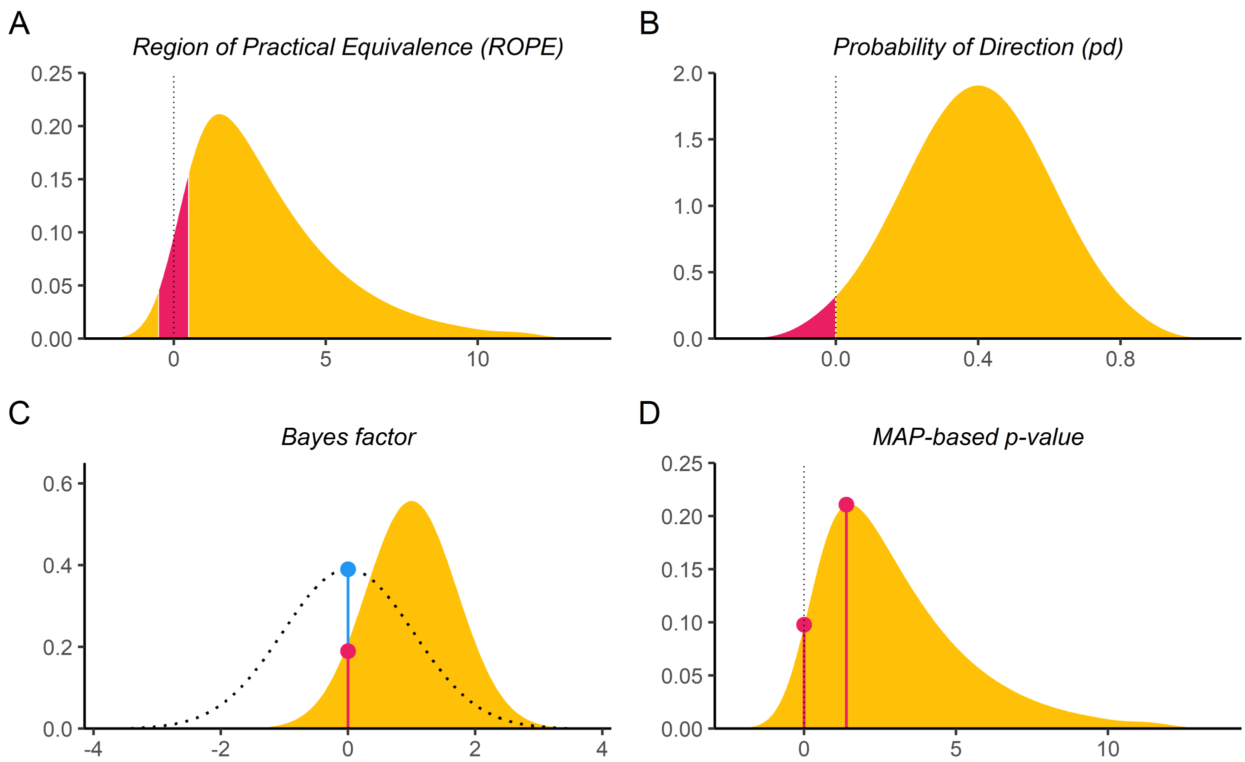

Figure 3: Posterior distributions (in yellow) of effects (the x-axis) with the illustration of indices of Null-Hypothesis Significance Testing: (A) The ROPE; (B) The Probability of Direction - pd; (C) The Savage-Dickey Bayes factor; (D) The MAP-based p-value.

Null-Hypothesis Significance Testing (NHST)

The Bayesian framework allows to neatly delineate and quantify different aspects of hypothesis testing, including effect *existence* and *significance*, and different indices have been developped to describe them.

ROPE and Test for Practical Equivalence

`rope()` computes the proportion of the HDI (default to the 89% HDI) of a posterior distribution that lies within a Region Of Practical Equivalence (the **ROPE**; see figure 3, panel A).

Statistically, the probability of a posterior distribution of being different from 0 does not make much sense (the probability of it being different from a single point being infinite). Therefore, the idea underlining ROPE is to let the user define an area around the null value enclosing values that are *equivalent to the null* value for practical purposes (Kruschke, 2018; Kruschke & Liddell, 2018). Nevertheless, in the absence of user-provided values, **bayestestR** will automatically find an appropriate range for the ROPE using the `rope_range()` function.

```
rope(distribution_normal(1000, mean = 1), range = c(-0.5, 0.5))
#> # Proportion of samples inside the ROPE [-0.50, 0.50]:
#>
#>  inside ROPE
#>      27.16 %
```

The proportion of HDI lying within this "null" region can be used as an decision criterion for "null-hypothesis" testing. Such **Test for Practical Equivalence**, implemented via `equivalence_test()`, is based on the

"HDI+ROPE decision rule" (Kruschke, 2018) to check whether parameter values should be accepted or rejected against an explicitly formulated "null hypothesis" (i.e., a ROPE). If the HDI is completely outside the ROPE, the "null hypothesis" for this parameter is "rejected". If the ROPE completely covers the HDI, i.e., all most credible values of a parameter are inside the ROPE, the null hypothesis is accepted. Else, it's undecided whether to accept or reject the null hypothesis.

```
library(rstanarm)
model <- stan_glm(mpg ~ wt + gear, data = mtcars)
equivalence_test(model)
#> # Test for Practical Equivalence
#>
#>   ROPE: [-0.60 0.60]
#>
#>    Parameter        H0 inside ROPE       89% HDI
#>  (Intercept)  Rejected      0.00 % [30.82 47.02]
#>           wt  Rejected      0.00 % [-6.63 -4.39]
#>         gear Undecided     52.54 % [-1.76  1.23]
```

Probability of Direction (*pd*)

`p_direction()` computes the **Probability of Direction** (*pd*, also known as the Maximum Probability of Effect - *MPE*). This index of effect existence varies between 50% and 100% and can be interpreted as the probability that a parameter (described by its posterior distribution) is strictly positive or negative (whichever is the most probable). It is mathematically defined as the proportion of the posterior distribution that is of the median's sign (see figure 3, panel B).

```
p_direction(distribution_normal(100, 0.4, 0.2))
#> # Probability of Direction (pd)
#>
#> pd = 98.00%
```

Bayes Factor

`bayesfactor_savagedickey()` computes the ratio between the density of a single value (typically the null) in two distributions. When these distributions are the prior and the posterior distributions, this ratio can be used to examine the degree by which the mass of the posterior distribution has shifted further away from or closer to the null value (relative to the prior distribution), thus indicating if the null value has become less or more likely given the observed data (see figure 3, panel C). The Savage-Dickey density ratio is also an approximation of a Bayes factor comparing the marginal likelihoods of the model against a model in which the tested parameter has been restricted to the point null (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010).

```
prior <- distribution_normal(1000, mean = 0, sd = 1)
posterior <- distribution_normal(1000, mean = 1, sd = 0.7)

bayesfactor_savagedickey(posterior, prior, direction = "two-sided", hypothesis = 0)
#> # Bayes Factor (Savage-Dickey density ratio)
#>
#>  Bayes Factor
#>          1.98
#> ---
#> Evidence Against Test Value:  0
```

MAP-based *p*-value

`p_map()` computes the odds that a parameter (described by its posterior distribution) has against the null hypothesis (*h0*) using Mills' *Objective Bayesian Hypothesis Testing* framework (Mills, 2018; Mills & Parent, 2014). It is mathematically based on the density at the Maximum A Priori (MAP) and corresponds to the density value at 0 divided by the density at the MAP estimate (see figure 3, panel D).

```
p_map(distribution_normal(1000, mean = 1))
#> # MAP-based p-value
#>
#> p (MAP) = 0.629
```

Licensing and Availability

bayestestR is licensed under the GNU General Public License (v3.0), with all source code stored at GitHub (https://github.com/easystats/bayestestR), with a corresponding issue tracker for bug-reporting and feature enhancements. In the spirit of honest and open science, we encourage requests/tips for fixes, feature updates, as well as general questions and concerns via direct interaction with contributors and developers.

Acknowledgments

bayestestR is part of the *easystats* ecosystem (relying on the **insight** package to access information contained in models; Lüdecke et al., 2019b), a collaborative project created to facilitate the usage of R. Thus, we would like to thank the council of masters of easystats, all other padawan contributors, as well as the users.

References

Andrews, M., & Baguley, T. (2013). Prior approval: The growth of bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 1–7.

Bürkner, P.-C., & others. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28.

Etz, A., & Vandekerckhove, J. (2016). A bayesian perspective on the reproducibility project: Psychology. *PloS One*, *11*(2), e0149794.

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2018). *Rstanarm: Bayesian applied regression modeling via Stan.* Retrieved from http://mc-stan.org/

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*(7), 293–300.

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2. ed). Amsterdam: Academic Press.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*(4), 722–752.

Kruschke, J. K., & Liddell, T. M. (2018). The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Lüdecke, D., Waggoner, P., Ben-Shachar, M. S., & Makowski, D. (2019a). *See: Visualisation toolbox for 'easystats' and extra geoms, themes and color palettes for 'ggplot2'*. Retrieved from https://easystats.github.io/see/

Lüdecke, D., Waggoner, P., & Makowski, D. (2019b). Insight: A unified interface to access information from model objects in r. *Journal of Open Source Software*, *4*(38), 1412. https://doi.org/10.21105/joss.01412

McElreath, R. (2018). *Statistical rethinking: A bayesian course with examples in r and stan*. Chapman; Hall/CRC.

Mills, J. A. (2018). Objective bayesian precise hypothesis testing. *University of Cincinnati*.

Mills, J. A., & Parent, O. (2014). Bayesian mcmc estimation. In *Handbook of regional science* (pp. 1571–1595). Springer.

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419. https://doi.org/10.1037/a0024377

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of bayes factors for common designs. *[Computer Software]*. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive Psychology*, *60*(3), 158–189.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … others. (2018). Bayesian inference for psychology. Part i: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57.