p-values and Bayes factors

Berenice López-Casal, Rocío Alcalá-Quintana[1], & Miguel A. García-Pérez
Universidad Complutense de Madrid

Bayes factors are gaining popularity as an alternative to p values computed under null hypothesis significance testing (NHST). While p values indicate the probability that departure of some magnitude from the null hypothesis occurs if the null is actually true, Bayes factors provide a measure of the strength of the evidence in favor/against the null.

For testing any hypothesis about distributional parameters, Bayes factors are computed from the exact same input that is used to compute a *p* value under NSHT plus, occasionally, the researcher's choice of a prior distribution. In these conditions, and given some choice on the prior, Bayes factors and *p* values must have a one-to-one correspondence that make them equivalent statistics that only express in a different manner what the data say about the hypotheses (Francis, 2017; García-Pérez, 2017).

This app illustrates the relation between *p* values and Bayes factors for various widely-used statistical tests of hypotheses.

t-test for a single mean or difference between two means with paired samples

This program simulates data, performs **two-sided Student's *t*-tests**, and reports the distributions of the corresponding NHST *p*-values and Bayes Factors for the null (BF01), where the latter are computed as described by Rouder et al. (2009).

Usage
Choose the type of test (either one sample or paired samples), the null hypothesis, and the scale parameter for the prior distribution to be used in the computation of Bayes factors. In order to generate the data, select the shape of the underlying distribution, along with the mean and standard deviation (for one sample t-tests) or the mean difference and standard deviation of the difference (for paired samples).

Output
The **scatter plot on the left** displays log BF01 against log *p*-values, showing a one-to-one correspondence between the two, and including the proportion of *p*-values smaller than 0.05 (bottom right, in red font) and BF01 smaller than 1 (top left, blue). These two points respectively mark the boundaries between rejecting versus not rejecting the null hypothesis and evidence opposing versus favoring the null. Note that multiple data points may overlap. The **histograms on the right** show the distribution of log Bayes Factors (top) and *p*-values (bottom), with tags indicating the proportion of $p < 0.05$, $p \geq 0.05$, BF01 < 1 and BF01 ≥ 1, as appropriate.

[1] To whom correspondence should be addressed (ralcala@psi.ucm.es)

Test for a single Pearson's correlation

This program performs **two-sided** tests for **Pearson's correlation** and reports the distributions of the corresponding NHST *p*-values and Bayes Factors for the null (BF01), where the latter is computed as described by Wetzels and Wagenmakers (2012).

Usage
You can change the true correlation used to generate the data, as well as the sample size, and the number of replicates. Pearson's correlation under the null hypothesis always equals zero, and, without loss of generality, the means and standard deviations of the normally distributed random variables equal 5 and 10, respectively.

Output
The **scatter plot on the left** displays log BF01 against log *p*-values, showing a one-to-one correspondence between the two, and including the proportion of *p*-values smaller than 0.05 (bottom right, in red font) and BF01 smaller than 1 (top left, blue). These two points respectively mark the boundaries between rejecting versus not rejecting the null hypothesis and evidence opposing versus favoring the null. Note that multiple data points may overlap. The **histograms on the right** show the distribution of log Bayes Factors (top) and *p*-values (bottom), with tags indicating the proportion of $p < 0.05$, $p \geq 0.05$, BF01 < 1 and BF01 ≥ 1, as appropriate.

Binomial test for a single proportion

This program performs **two-sided** tests for a single proportion in a (finite) **Bernoulli process** and reports the distributions of the corresponding NHST *p*-values and Bayes Factors for the null (BF01), where the latter is computed as described by Wagenmakers et al. (2010).

Usage
You can change the true probability of success used to generate the data, the probability of success under the null hypothesis, and the shape parameters (*a* and *b*) of the Beta prior distribution[2] along with the number of trials and replicates.

Output
The **scatter plot on the left** displays log BF01 against log *p*-values, showing a one-to-one correspondence between the two, and including the proportion of *p*-values smaller than 0.05 (bottom right, in red font) and BF01 smaller than 1 (top left, blue). These two points respectively mark the boundaries between rejecting versus not rejecting the null hypothesis and evidence opposing versus favoring the null. Note that multiple data points may overlap. The **histograms on the right** show the distribution of log Bayes Factors (top) and *p*-values (bottom), with tags indicating the proportion of $p < 0.05$, $p \geq 0.05$, BF01 < 1 and BF01 ≥ 1, as appropriate.

Note: Try and change the shape parameters for the prior Beta distribution (e.g., Beta(3,1)) while keeping at 0.5 the true probability of success and the probability of success under H0, and watch how the scatterplot changes into a two-limbed curve whenever the prior is asymmetric. The lower limb

[2] Note that Beta(1,1) is equivalent to a (non-informative) uniform prior distribution.

includes all the replicates where the evidence (i.e., data) gathered lies to the left of H0, whilst the upper limb corresponds to evidence that lies to the right of such point. This two-limbed curve apparently breaks the one-to-one correspondence between *p* values and Bayes factors, but it only says that the relation varies according to whether observed data that are equally likely under the null fall on the side of the (biased) prior or on the other.

References

Francis, G. (2017) Equivalent statistics and data interpretation. *Behavior Research Methods*, in press. Advance online publication DOI 10.3758/s13428-016-0812-3

García-Pérez, M. A. (2017). Thou shalt not bear false witness against null hypothesis significance testing. *Educational and Psychological Measurement*, in press. Advance online publication DOI 10.1177/0013164416668232

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225-237.

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.

Wetzels, R., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057-1064.