

# Similarities and differences in behaviour of the BIC based and likelihood ratio (LR) based Bayes Factor outputs under simple RCT analysis to two-sided p values from two sample t-tests for fixed sample sizes.

*John Martin*

*3/27/2020*

## Executive Summary

For simple Randomised Control Trial (RCT) analysis using a nested model approach the sample size normalised BIC based Bayes factor estimates is identical to a simple Likelihood Ratio (LR) based Bayes Factor behaviour

$$\frac{1}{\sqrt{N}} \frac{P_{BIC}(D|H0)}{P_{BIC}(D|H1)} = \exp(-LR/2) \equiv \frac{\mathcal{L}(H0)}{\mathcal{L}(H1)}.$$

## Introduction

There is continuing disagreement [1,2] about the correlation between

- $p_{value}$  the two sample t-test probability of the observed data given the Null Hypothesis and
- $P(H0|D)$  the Bayesian posterior probability of the Null Hypothesis given the observed data.

In this paper, the normalised BIC based estimates Bayes Factor  $\frac{1}{\sqrt{N}} \frac{P_{BIC}(D|H0)}{P_{BIC}(D|H1)}$  described in [3] which has a strong contour line relationship with  $p_{value}$  for simple RCT analysis under repeated sampling for fixed sample size is illustrated to closely approach simple likelihood based Bayes Factor behaviour.

Firstly, the contrasting behaviour for the relationship of BIC based Bayes Factor  $P_{BIC}(D|H0)$  & likelihood based Bayes Factor  $P_{LR}(D|H0)$  outputs with  $p_{value}$  is shown for A/A & A/B conditions for RCT analysis under repeated sampling for fixed sample sizes.

Then the close behaviour for the Bayes Factor estimates  $\frac{1}{\sqrt{N}} \frac{P_{BIC}(D|H0)}{P_{BIC}(D|H1)}$  &  $\frac{P_{LR}(D|H0)}{P_{LR}(D|H1)}$  is shown.

## BIC based Bayes Factor estimation

Masson [4] demonstrated how to calculate estimates of relative posterior probabilities using

$$\frac{P(H0|D)}{P(H1|D)} = \frac{P(D|H0)}{P(D|H1)} \cdot \frac{P(H0)}{P(H1)} \quad (1)$$

$$= BF \cdot \frac{P(H0)}{P(H1)} \quad (2)$$

where BF is the Bayes Factor of the ratio of the likelihoods of the data given H0 and H1 respectively, by using Bayesian Information Criteria (BIC) model fit calculations

$$BIC = -2\ln(L) + k\ln(n) \quad (3)$$

where  $L$  is the maximum likelihood of the fitted model,  $k$  is the number of free model parameters and  $n$  is the sample size.

Explicitly, the BIC based estimate [4] for the Bayes Factor component in (2), is given by

$$\frac{P(H0|D)}{P(H1|D)} \approx \frac{P_{BIC}(D|H0)}{P_{BIC}(D|H1)} \cdot \frac{P(H0)}{P(H1)} \quad (4)$$

$$\approx e^{\frac{(\Delta BIC)}{2}} \cdot \frac{P(H0)}{P(H1)} \quad (5)$$

$$\rightarrow e^{\frac{(\Delta BIC)}{2}} \quad \text{as } \frac{P(H0)}{P(H1)} \rightarrow 1 \quad (6)$$

where

$$\Delta BIC = BIC_{H1} - BIC_{H0} \quad (7)$$

As illustrated in [3], a useful normalised version of the above estimator to compare to  $p_{values}$  is

$$\text{norm}P_{BIC} = \frac{1}{\sqrt{N}} \frac{P_{BIC}(D|H0)}{P_{BIC}(D|H1)} \quad (8)$$

## Likelihood Ratio based Bayes Factor estimation

The above approach by Masson can be replicated using a Likelihood Ratio (LR) based estimate for the Bayes Factor where for a nested set model

$$LR = -2 \cdot \ln\left(\frac{\mathcal{L}(H0)}{\mathcal{L}(H1)}\right) \quad (9)$$

where  $\mathcal{L}(H0)$ ,  $\mathcal{L}(H1)$  are the maximum likelihoods of the fitted models.

Explicitly, the LR based estimate for the Bayes Factor component in (2), could be estimated by

$$\frac{P(H0|D)}{P(H1|D)} \approx \frac{P_{LR}(D|H0)}{P_{LR}(D|H1)} \cdot \frac{P(H0)}{P(H1)} \quad (10)$$

$$\approx e^{\frac{-(LR)}{2}} \cdot \frac{P(H0)}{P(H1)} \quad (11)$$

$$\rightarrow e^{\frac{-(LR)}{2}} \quad \text{as } \frac{P(H0)}{P(H1)} \rightarrow 1 \quad (12)$$

where

$$e^{\frac{-(LR)}{2}} = e^{-1 \cdot -2 \cdot \ln\left(\frac{\mathcal{L}(H0)}{\mathcal{L}(H1)}\right) \cdot \frac{1}{2}} \quad (13)$$

$$= e^{\ln\left(\frac{\mathcal{L}(H0)}{\mathcal{L}(H1)}\right)} \quad (14)$$

$$= \frac{\mathcal{L}(H0)}{\mathcal{L}(H1)} \quad (15)$$

Importantly, it is already demonstrated that the frequentist based null hypothesis significance test (NHST) method calculated via the probability of the observed t statistic under a two-sided null hypothesis can be related to a LR test [5,6].

Explicitly, the estimator for the numerator of the Likelihood based Bayes Factor component in (2), is given by

$$P_{LR}(D|H_0) = \frac{\frac{\mathcal{L}(H_0)}{\mathcal{L}(H_1)}}{\frac{\mathcal{L}(H_0)}{\mathcal{L}(H_1)} + 1} \quad (16)$$

## Relationship between BIC based and Likelihood Ratio based Bayes Factors

Using equations (3), (6) and (15) it can be seen for simple RCT analysis

$$\frac{(\Delta BIC)}{2} = \frac{(BIC_{H_1} - BIC_{H_0})}{2} \quad (17)$$

$$= \frac{(-2\ln(\mathcal{L}_{H_1}) + 1 \cdot \ln(n) + 2\ln(\mathcal{L}_{H_0}))}{2} \quad (18)$$

$$= \frac{(-2\ln(\frac{\mathcal{L}_{H_1}}{\mathcal{L}_{H_0}}) + 1 \cdot \ln(n))}{2} \quad (19)$$

$$= \frac{(-2\ln(\frac{\mathcal{L}_{H_1}}{\mathcal{L}_{H_0}}) - 2 \cdot \ln(\frac{1}{n^{\frac{1}{2}}}))}{2} \quad (20)$$

$$= \frac{(-2\ln(\frac{\mathcal{L}_{H_1}}{\mathcal{L}_{H_0}} \cdot \frac{1}{n^{\frac{1}{2}}}))}{2} \quad (21)$$

$$= -\ln(\frac{\mathcal{L}_{H_1}}{\mathcal{L}_{H_0}} \cdot \frac{1}{n^{\frac{1}{2}}}) \quad (22)$$

$$(23)$$

$$\therefore e^{\frac{(\Delta BIC)}{2}} = e^{-\ln(\frac{\mathcal{L}_{H_1}}{\mathcal{L}_{H_0}} \cdot \frac{1}{n^{\frac{1}{2}}})} \quad (24)$$

$$= \frac{\sqrt{n}}{1} \cdot \frac{\mathcal{L}_{H_0}}{\mathcal{L}_{H_1}} \quad (25)$$

$$\equiv \sqrt{n} \cdot e^{-LR/2} \quad (26)$$

and hence for simple RCT analysis

$$\text{norm}P_{BIC} = \frac{1}{\sqrt{n}} \frac{P_{BIC}(D|H_0)}{P_{BIC}(D|H_1)} \quad (27)$$

$$\equiv \frac{\mathcal{L}(H_0)}{\mathcal{L}(H_1)} \quad (28)$$

which is bounded by (0,1) and has a known direct relationship to NHST *p-values*.

## Likelihood Ratio (and normalised BIC based) Bayes Factor vs $p_{value}$ behaviour for different sample sizes

As in [3] the following figure 1 shows the pseudo ROC curve relationship between  $BF_{LR} \equiv \frac{BF_{BIC}}{\sqrt{N}}$  and  $p_{value}$ . On the figure are thresholds indicating strength of evidence for H1 including the p-value based significance regions, the Bayesian inference significance regions using Jeffreys [7] categories and the large N intersection of the p-value based and BIC based Bayesian thresholds (or simply the intersection of p-value and Likelihood Ratio based Bayes Factors) for simple RCT analysis.

An interpretation of the Likelihood Ratio based Bayes Factor is that it calculates the relative maximum likelihood performance of the null model (H0) compared to the alternative model (H1) in explaining the dependent data variable behaviour. Likewise using equations (28) and (3) the BIC based Bayes Factor calculates the relative maximum likelihood performance of the null model (H0) compared to a penalised alternative model (H1). With the conclusion that the penalty term for the alternative model (H1) is the factor that gives rise to Lindley's Paradox situations between BIC Bayes Factor estimates and NHST  $p_{value}$  estimates (given there is a relative closeness between NHST  $p_{value}$  cutoffs and Likelihood Ratio estimate significance cutoffs as echoed in figure 1.)

# (BIC based Bayes Factor)/sqrt(N) vs p\_value

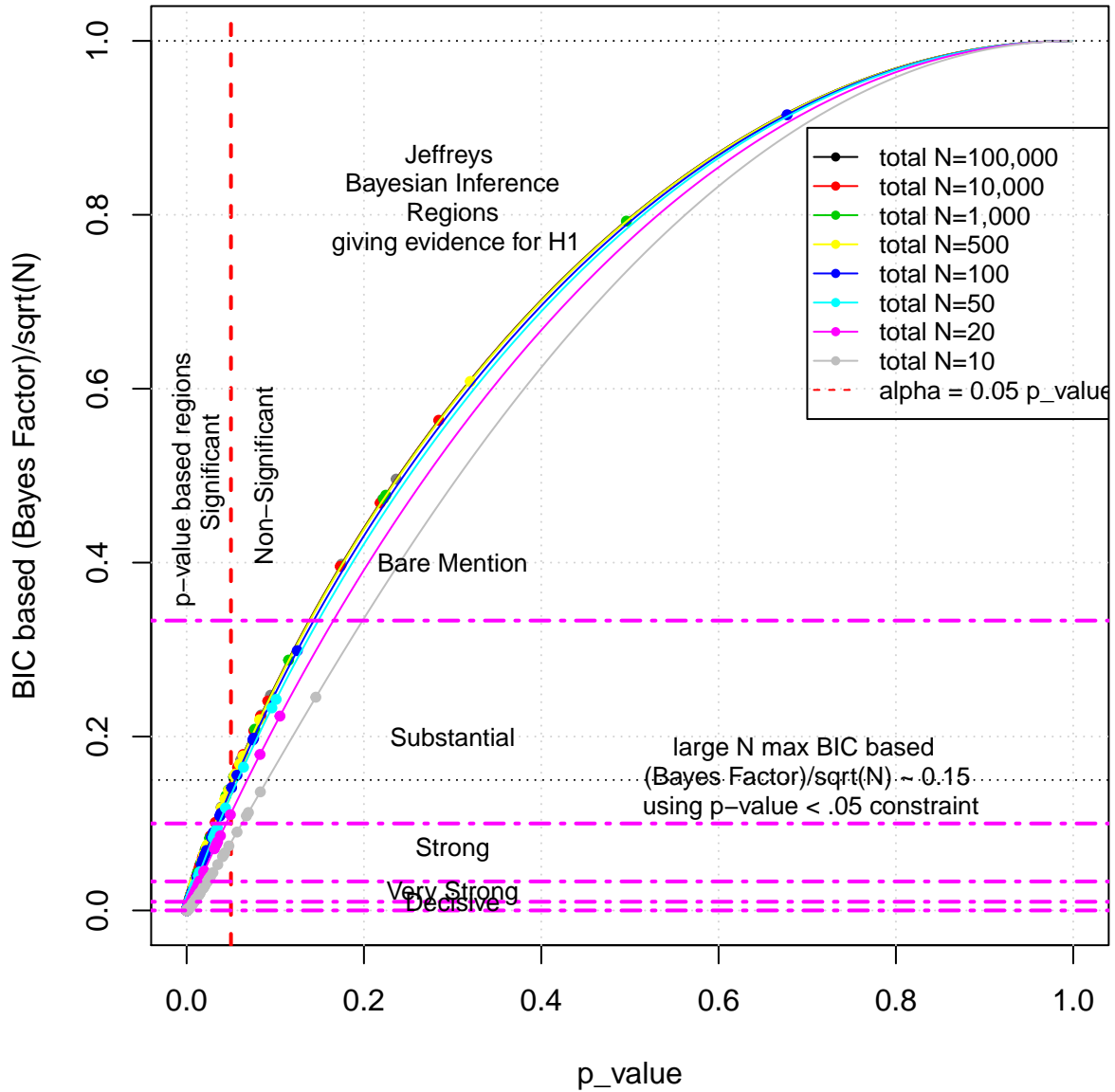


Figure 1. ROC curve behaviour of LR based Bayes factor ( $\equiv$  normalised BIC based  $(BF/\sqrt{N})$ ) is shown as a function of  $p_{value}$  for different sample sizes. A overlay mapping of the Bayesian based strength of evidence regions (horizontal dot-dash lines) in favour of  $H_1$  and the  $\alpha = 0.05$  based cutoffs to reject  $H_0$  (vertical dashed line) are also shown. The data points are from simple RCT analysis when  $P(H_1) = 1$ , the statistical power is  $\sim 95\%$ , and for clarity only 100 repeated samples are shown. The contour lines behind the data points were obtained from figure 2 using  $P(H_0) = 1$  to indicate the contour line positions are independent of the RCT treatment effect.

## Empirical behaviour of $P_{LR}(D|H0)$ & $P_{BIC}(D|H0)$ vs $p_{value}$ for simple RCT analysis under repeated sampling for different fixed sample sizes

Figures 2 & 3, show the significant difference in  $P_{LR}(D|H0)$  &  $P_{BIC}(D|H0)$  outputs for simple RCT analysis. The  $P_{LR}(D|H0)$  versus  $p_{value}$  behaviour does not have the Lindley Paradox issues that is exhibited for  $P_{BIC}(D|H0)$  versus  $p_{value}$  behaviour.

It is informative that when  $p_{value} = 1$ , the numerator and denominator of the Likelihood Ratio based Bayes Factor are  $P_{LR}(D|H0) = 0.5$  &  $P_{LR}(D|H1) = 0.5$  indicating that both (unpenalised) models are equally efficient at explaining the data behaviour. Under these conditions the role of the prior odds becomes the determining factor in estimating the posterior odds.

Under A/A conditions ie.  $P(H1) = 0$

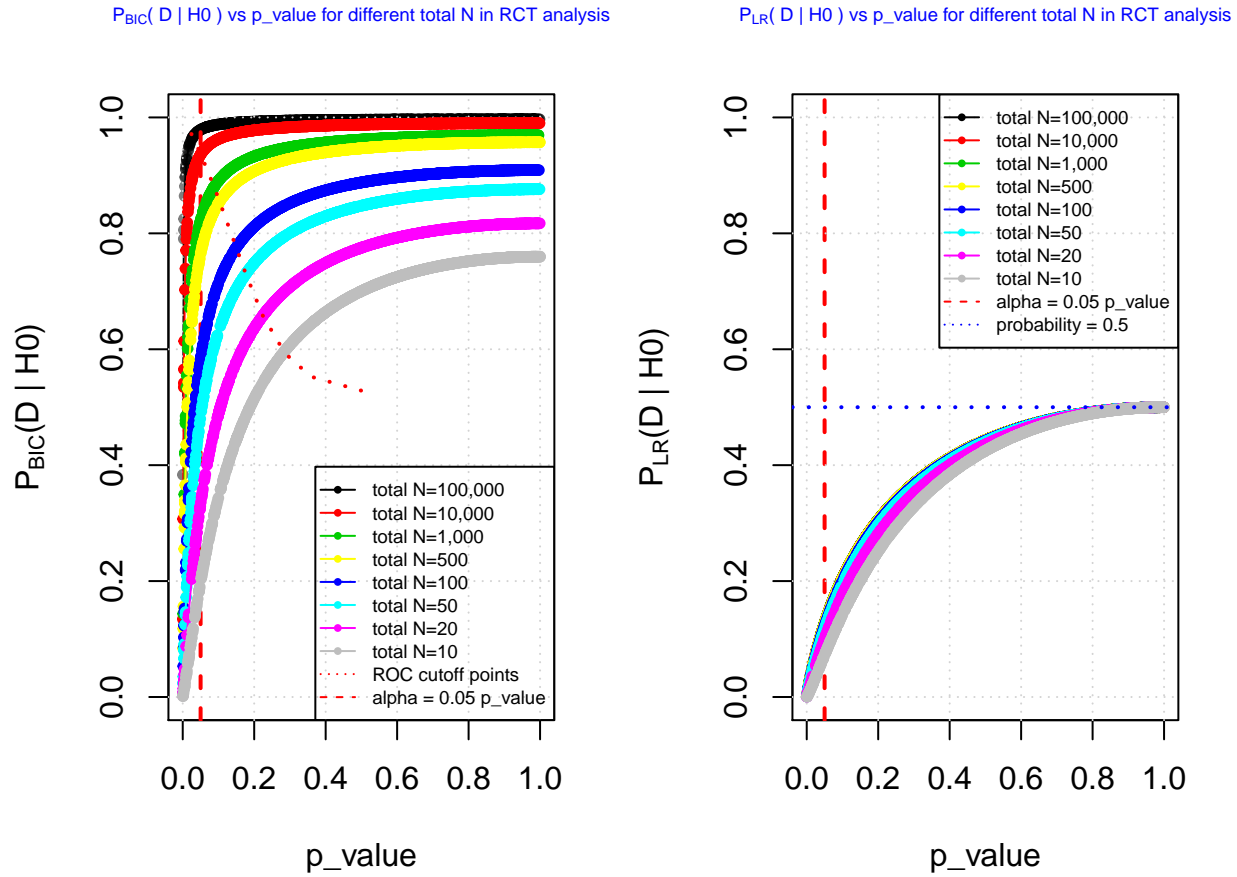


Figure 2.  $P_{BIC}(D|H0)$  as a function of  $p$  value for different sample sizes under simple RCT analysis when  $P(H0) = 1$ .

Under A/B conditions ie.  $P(H1) = 1$

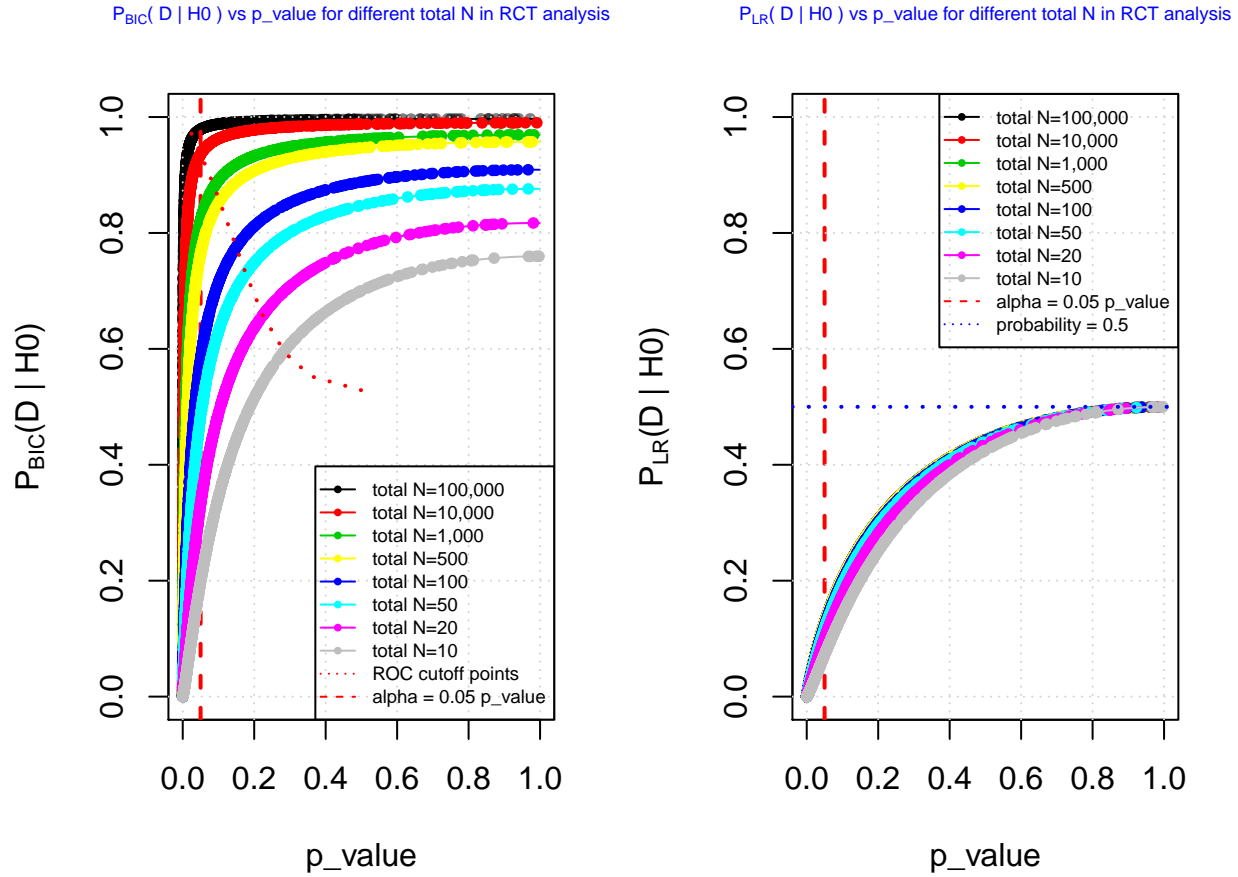


Figure 3.  $P_{BIC}(D|H0)$  as a function of  $p$  value for different sample sizes under simple RCT analysis when  $P(H1) = 1$ . As evidence, of fixed contour line behaviour the fitted lines on the graph are from  $P(H1) = 0$  data and these lines fit the  $P(H1) = 1$  data points very well.

## Conclusions

By normalising the BIC based Bayes Factor estimates for simple RCT analysis, the values are identical to a Likelihood Ratio based Bayes Factor approach and the Lindley Paradox conflict is not relevant.

## References

1. Trafimow, D., & Rice, S. (2009). A test of the null hypothesis significance testing procedure correlation argument. The Journal of General Psychology, 136, 261-270.
2. Lakens, D. (2015) <https://daniellakens.blogspot.com/2015/11/the-relation-between-p-values-and.html>
3. Martin, J.P.D. (2020) The strong contour line relationship between two-sided  $p$  values of simple RCT two sample t-tests and BIC based Bayes factor  $P_{BIC}(D|H0)$  outputs for fixed sample sizes. DOI 10.6084/m9.figshare.12003945

4. Masson, M. E. J. (2011) A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behav Res* 43:679–690 DOI 10.3758/s13428-010-0049-5
5. Neyman, J., Pearson, E. S. (1933) On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London A*, 231 (694–706) doi:10.1098/rsta.1933.0009,
6. Francis, G. (2017) Equivalent statistics and data interpretation. *Behav Res* 49:1524–1538 <https://doi.org/10.3758/s13428-016-0812-3>
7. Jeffreys H. (1961) *Theory of Probability*. Oxford, UK: Oxford Univ. Press. 3rd ed.