Jeffreys–Lindley Paradox

From Basic to Theoretical

Zhengxiao Wei
Dr. Nathoo's Lab

University of Victoria

October 1, 2025

Paradoxes

Quine (1976) distinguished three types of paradoxes.

- Veridical paradoxes (surprising yet true)

  Hilbert's hotel, Monty Hall problem, Simpson's paradox

- Falsidical paradoxes (seem sound but hinge on a mistake)

  Zeno's paradoxes

- Antinomies (accepted principles yield contradiction)

  Russell's paradox, temporal paradox, Ship of Theseus, Schrödinger's cat

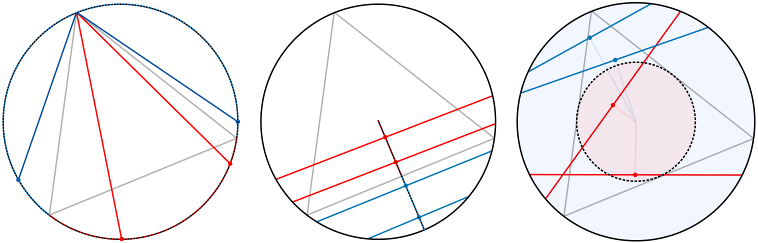
▷ *Better seen as dilemmas*

  Fermi paradox, Newcomb's problem, trolley problem

A Falsidical Paradox

The Bertrand paradox and the maximum ignorance principle:

What is the probability that a random chord of a circle is longer than the side of an equilateral triangle inscribed within that circle?



Source: Wikipedia

1/3, 1/2, and 1/4 are all correct (or incorrect) depending on the question being asked (i.e., random endpoints, random radius, and random midpoint).

Two Views of Probability

Frequentist

- Probability is the long-run frequency.
- $p(y; \boldsymbol{\theta})$ is a sampling distribution:

  a function of y with $\boldsymbol{\theta}$ fixed.

  It assumes repeated sampling from the population to estimate the true value of an underlying parameter.

$\bar{Y}_n \xrightarrow{P} \mu$ as $n \to \infty$

Bayesian

- Probability reflects updated beliefs.
- $p(y \mid \boldsymbol{\theta})$ is a likelihood:

  a function of $\boldsymbol{\theta}$ with y fixed.

  Assuming a proper prior and i.i.d. data, the posterior becomes more concentrated as the sample size increases.

$p(\boldsymbol{\theta} \mid y) \propto p(y \mid \boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta})$

A Veridical Paradox

The Jeffreys–Lindley paradox refers to the disagreement between the frequentist and Bayesian approaches to a hypothesis testing problem, $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $\mathcal{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

A p-value may *reject* the null hypothesis, while the Bayes factor may *support* it, especially with a small effect size, diffuse prior, and large sample size.

$$p\text{-value} := \mathbb{P}(|T| \geqslant |t(\boldsymbol{y})|; \mathcal{H}_0) < .05 \tag{1}$$

$$BF_{01} := \frac{p(\boldsymbol{y} \mid \mathcal{H}_0)}{p(\boldsymbol{y} \mid \mathcal{H}_1)} = \frac{\mathbb{P}(\mathcal{H}_0 \mid \boldsymbol{y})}{\mathbb{P}(\mathcal{H}_1 \mid \boldsymbol{y})} \Big/ \frac{\mathbb{P}(\mathcal{H}_0)}{\mathbb{P}(\mathcal{H}_1)} \tag{2}$$

$$= \frac{\int p(\boldsymbol{y} \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\psi}) \cdot \pi_0(\boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\psi}}{\iint p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\psi}) \cdot \pi_1(\boldsymbol{\psi} \mid \boldsymbol{\theta}) \cdot \pi_1(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\psi} \, \mathrm{d}\boldsymbol{\theta}} \tag{3}$$

$$= \frac{p(\boldsymbol{\theta} = \boldsymbol{\theta}_0 \mid \boldsymbol{y}, \mathcal{H}_1)}{p(\boldsymbol{\theta} = \boldsymbol{\theta}_0 \mid \mathcal{H}_1)} > 3 \tag{4}$$

Example: $Y \sim B(n, \theta)$

Source BayesFactor

```r
k <- 49581; N <- 98451; k/N # 0.5036109

binom.test(k, N)
  # Exact binomial test
  # --------------
  # data:  k and N
  # number of successes = 49581, number of trials = 98451, p-value = 0.02365
  # alternative hypothesis: true probability of success is not equal to 0.5
  # 95 percent confidence interval:
  #  0.5004826 0.5067390
  # sample estimates:
  # probability of success
  #              0.5036109

1 / BayesFactor::proportionBF(k, N, .5)  # BF[01] = 1 / BF[10]
  # Bayes factor analysis
  # --------------
  # [1] Null, p=0.5 : 9.607493 ±0%
  #
  # Against denominator:
  #   Alternative, p0 = 0.5, r = 0.5, p =/= p0
  # ---
  # Bayes factor type: BFproportion, logistic

dbeta(.5, k+1, N-k+1) / dbeta(.5, 1, 1) # 19.21139, Savage-Dickey density ratio
```

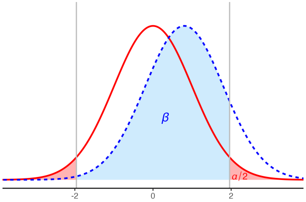
Hypothesis Testing

p-Value

- Absense of evidence
  $\neq$ Evidence of absence

  p-value $\neq \mathbb{P}(\mathcal{H}_0)$

- The frequentist finds that $\mathcal{H}_0$ is a poor explanation for the data (without reference to $\mathcal{H}_1$).



Bayes Factor

- Model selection
  $\neq$ Model comparison

  All models are wrong,
  but some are useful (Box, 1976).

- The Bayesian finds that $\mathcal{H}_0$ is a relatively better explanation for the data than $\mathcal{H}_1$.

  Evidence is relative.

What if a model is misspecified?

Conversely, the Bayes factor may favor $\mathcal{H}_1$, while the p-value may fail to reject $\mathcal{H}_0$.

(Wei, Nathoo, & Masson, 2025).

- Clustered data

  e.g., the weight of the jth piglet in the randomly selected litter i

$$\begin{aligned} &\mathcal{M}_1: \quad y_{ij} = \mu + b_i + \epsilon_{ij} \\ \text{versus} \quad &\mathcal{M}_0: \quad y_{ij} = \mu + \epsilon_{ij} \end{aligned}$$

- One-way within-subject data

  e.g., the response time of the ith participant for the jth Stroop task

Same mixed model but different designs

$$\begin{aligned} &\mathcal{M}_1: \quad y_{ij} = \mu + b_i + t_j + \epsilon_{ij} \\ \text{versus} \quad &\mathcal{M}_0: \quad y_{ij} = \mu + b_i + \epsilon_{ij} \end{aligned}$$

Lindley (1957) and Bartlett (1957)

Bayesian z-test.

Data: $Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta,\, \sigma^2)$ for $i = 1, \cdots, n$

$\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta \neq \theta_0$

Prior: $\theta \sim \mathcal{N}(\theta_0,\, \sigma^2)$

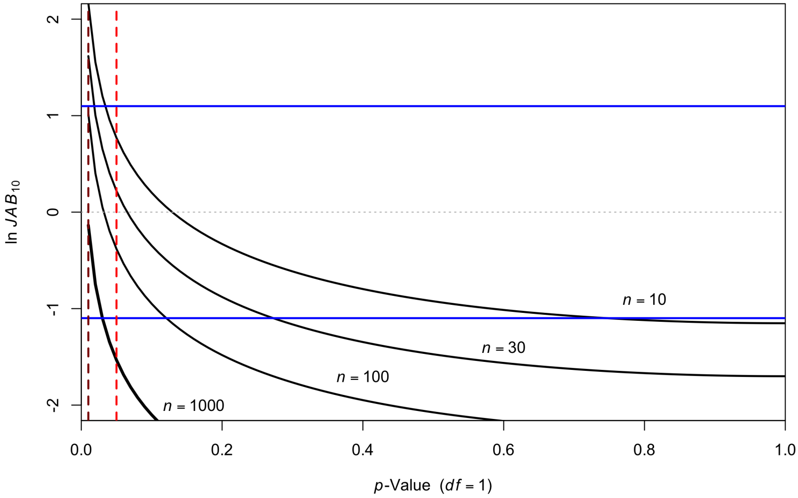
$$BF_{01} = (1+n)^{\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{n}{1+n} \cdot z^2\right\} \xrightarrow{P} \infty \quad \text{as} \quad n \to \infty, \tag{5}$$

assuming a fixed $z^2 = n(\bar{Y}_n - \theta_0)^2/\sigma^2$.

Jeffreys (1936)

Assume a unit-information prior in the Wald test.

$$JAB_{01} = \sqrt{n} \cdot \exp\left\{-\frac{1}{2}W\right\} \tag{6}$$



Jeffreys–Zellner–Siow Bayes Factor

Bayesian one-sample t-test.

Data: $Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\sigma\delta,\, \sigma^2)$ for $i = 1, \cdots, n$

$\mathcal{H}_0 : \delta = 0$ versus $\mathcal{H}_1 : \delta \neq 0$

Priors: $\pi(\sigma) \propto 1\,/\,\sigma$ and $\delta \sim \mathrm{Cauchy}(0,\, h)$

$$\textit{JZS-BF}_{10} = (2\pi)^{-\frac{1}{2}} h \left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}} \int_0^\infty (1+ng)^{-\frac{1}{2}} \left(1 + \frac{t^2}{(1+ng)\nu}\right)^{-\frac{\nu+1}{2}} g^{-\frac{3}{2}} e^{-\frac{h^2}{2g}}\, \mathrm{d}g, \qquad (7)$$

where $\nu = n - 1$ and (default) $h = \sqrt{2}/2$.

Question: Since p-values and Bayes factors are equated, does this imply that we can set a threshold for p-values that ‘accepts’ the null hypothesis?

Takeaways

Basic intuition

In essence, the apparent disagreement between the methods is not a disagreement at all, but rather two different statements about how the hypotheses relate to the data.

Theoretical results

$$BF_{01} \gtrsim -e \cdot p \ln p \quad \text{for } p < 1\,/\,e$$

$$JAB_{01} = \sqrt{n} \cdot \exp\left\{-\frac{1}{2}\,W\right\}$$

References

Bartlett, M. S. (1957). A comment on D. V. Lindley's statistical paradox. *Biometrika*, *44*, 533–534.
Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*, 791–799.
Held, L., & Ott, M. (2018). On p-values and Bayes factors. *Annual Review of Statistics and Its Application*, *5*, 393–419.
Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, *31*, 203–222.
Jeffreys, H. (1936). Further significance tests. *Mathematical Proceedings of the Cambridge Philosophical Society*, *32* , 416–445.
Lindley, D. V. (1957). A statistical paradox. *Biometrika*, *44*, 187–192.
Ly, A., Verhagen, J., & Wagenmakers, E.-J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32.
Morey, R. D., Romeijn, J. W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, *72*, 6–18.
Quine, W. V. O. (1976). *The ways of paradox and other essays* (2nd ed.). Harvard University Press.
Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. *Psychological Methods*, *22*, 304–321.
Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779–804.
Wagenmakers, E.-J. (2022). Approximate objective Bayes factors from p-values and sample size: The $3p\sqrt{n}$ rule. *PsyArXiv*. 1–50.
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*, 158–189.
Wagenmakers, E.-J., & Ly, A. (2023). History and nature of the Jeffreys–Lindley paradox. *Archive for History of Exact Sciences*, *77*, 25–72.
Wei, Z., Nathoo, F. S., & Masson, M. E. J. (2025). *Evaluating the Bayes factor under model misspecification*. Manuscript submitted for publication.

Thank you.

Contact: zhengxiao@uvic.ca