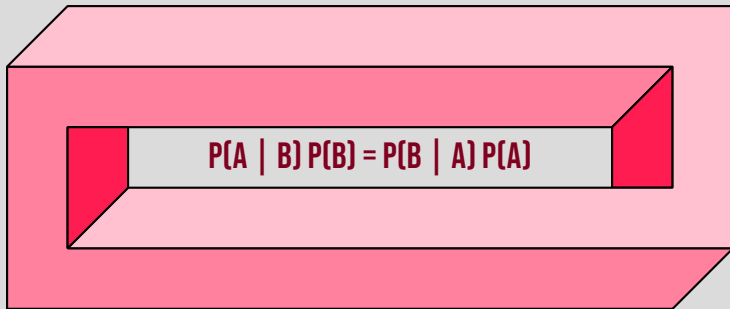


SYMMETRIES AND PARADOXES IN BAYESIAN STATISTICS



THE 1ST INTERNATIONAL ONLINE CONFERENCE ON SYMMETRY | OCTOBER 2026

ANDREW FOWLIE | XJTLU | ANDREW.FOWLIE@XJTLU.EDU.CN | ANDREWFOWLIE.GITHUB.IO

BASED ON: DOI:10.1080/03610926.2020.1866206

CONTENTS

1. PARADOXES, SYMMETRY, AND PRIORS

2. THE LIMITS OF SYMMETRY

3. THE JEFFREYS-LINDLEY PARADOX

4. SUMMARY

SECTION 1

PARADOXES, SYMMETRY, AND PRIORS

WHAT IS A PARADOX?

- ▶ A paradox is often not a contradiction in the rules of logic or probability
- ▶ Paradoxes **expose** assumptions that were left implicit
- ▶ Paradoxes help us **understand** foundations of a subject
- ▶ E.g., from physics, the EPR paradox in quantum mechanics; the ladder paradox in special relativity; or Maxwell's demon in statistical mechanics
- ▶ Paradoxes play the same role in statistics

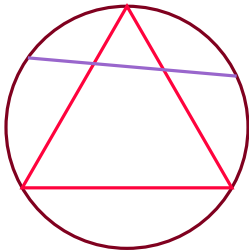
WHEN PROBABILITY DEFIES INTUITION

- ▶ **Monty Hall:** Under the usual rules, switching doors wins with probability $2/3$. The host's knowledge and behaviour matter (Selvin, 1975)
- ▶ **Simpson's paradox:** A trend seen within each subgroup can reverse when the groups are combined (Simpson, 1951)
- ▶ **Bertrand's chord paradox:** Different ways of choosing a “random” chord give different answers (Bertrand, 1888)
- ▶ **Jeffreys-Lindley paradox:** Differences between Bayesian and frequentist conclusions can grow as the size of dataset increases (Jeffreys, 1939; Lindley, 1957)

Paradoxes often reveal a **hidden assumption or misconception** about conditioning, grouping, what question we are asking, or how randomness is generated

BERTRAND'S CHORD PARADOX

Inscribe an equilateral triangle in a circle. What is the probability that a “random” chord is longer than a side of the triangle (Bertrand, 1888)?



Which **symmetry**, if any, should determine the prior? Is the problem well-posed (Jaynes, 1973)? How much can symmetry help us?

SECTION 2

THE LIMITS OF SYMMETRY

HOW MUCH CAN SYMMETRY TELL US?

Suppose we agree that the prior should be invariant under **rotations** of the circle:

How the chord is generated	Probability
Choose two endpoints uniformly on the circumference	1/3
Choose a radius, then a midpoint uniformly along it	1/2
Choose the midpoint uniformly over the disk	1/4

- ▶ Each chord-generation procedure **respects rotations of the circle**
- ▶ Thus, rotational symmetry alone **cannot** select a unique distribution over chords

SYMMETRIES & “FLATNESS”

Suppose a parameter θ is replaced by $\phi = g(\theta)$. The prior density transforms as

$$\pi_{\phi}(\phi) = \pi_{\theta}(g^{-1}(\phi)) \left| \frac{dg^{-1}(\phi)}{d\phi} \right|$$

- ▶ A density that is constant in one coordinate need not be constant in another
- ▶ “Flat” therefore depends on the parametrisation and “flat” doesn’t specify a prior
- ▶ Invariance can guide prior construction if the relevant symmetry is stated

Invariance and symmetries are **modelling principles** — not a synonym for having no assumptions or equivalent to “flat”

IMPROPER PRIORS: USEFUL, BUT DELICATE

- ▶ Invariant measures associated with non-compact symmetry groups are often **improper**
- ▶ An improper prior cannot be normalized to integrate to one; e.g., $\pi(\theta) \propto 1/\theta$
- ▶ It may still produce a proper posterior in some estimation problems.
- ▶ But its arbitrary normalising constant generally prevents a well-defined marginal likelihood
- ▶ Bayes factors compare marginal likelihoods, so improper priors require particular care in model comparison

Tension — a prior motivated by symmetry may be improper—and therefore unsuitable for a Bayes factor

SECTION 3

THE JEFFREYS-LINDLEY PARADOX

A QUESTION ABOUT EVIDENCE

- ▶ What should count as evidence for or against a hypothesis?
- ▶ **Bayesians** quantify evidence using probability theory and the data at hand
- ▶ **Frequentists** consider what would happen in repeated sampling of data, usually assuming the null hypothesis is true
- ▶ **Fisherians** interpret frequentist results as evidence
- ▶ Their answers may behave differently as the sample size changes

Jeffreys–Lindley Paradox — Differences between frameworks may **grow** as sample size increases

A SIMPLE TESTING PROBLEM

- ▶ Suppose n samples are independent and identically normally distributed with known variance σ^2 but unknown mean μ :

$$x_i \sim \mathcal{N}(\mu, \sigma^2) \quad \text{for } i = 1, \dots, n$$

- ▶ We test between two hypotheses: the null, H_0 , that $\mu = 0$; and the alternative, H_1 , that $\mu \neq 0$
- ▶ The data can be summarized by the sufficient statistic

$$t \equiv \frac{\sqrt{n} \bar{x}}{\sigma}$$

- ▶ Under the null, the test statistic follows a standard normal, $t \sim \mathcal{N}(0, 1)$
- ▶ After observing t , our Bayesian uses a **Bayes factor**; our **frequentist** uses a ***p*-value**

THE P-VALUE

Under the null hypothesis, how surprising are data at least as extreme as those observed?

$$p = \Pr_0(|t| \geq |t_{\text{obs}}|)$$

- ▶ Measure of disagreement between data and null
- ▶ Does not give the probability of the null

THE BAYES FACTOR

How well does the observed data t support one hypothesis relative to the other?

$$B_{01} = \frac{\Pr(t_{\text{obs}} | H_0)}{\Pr(t_{\text{obs}} | H_1)}$$

- ▶ Compares likelihoods of data under null and alternative
- ▶ Requires a prior for unknown μ under the alternative

p measures extremeness under the null; B compares how well the hypotheses predict the data

THE JEFFREYS-LINDLEY PARADOX

Consider keeping t fixed while increasing n :

- ▶ The frequentist p -value stays fixed e.g. at 1% — it **only depends on t**
- ▶ The Bayes factor can increasingly favour the null — **it depends on t and n**
- ▶ Thus, the two approaches can give conflicting assessments of the same data, and the conflict increases with n

What assumptions or symmetries would be required to avoid the growing discrepancy between p and B (Fowlie, 2020)?

A SYMMETRY VIEWPOINT

Under the alternative model, the test statistic follows,

$$t \mid \theta \sim \mathcal{N}(\theta, 1) \quad \text{for} \quad \theta = \frac{\sqrt{n} \mu}{\sigma},$$

The predictions are naturally expressed in terms of the rescaled parameter θ , while the underlying prior is specified on the unknown μ

- ▶ Changing n changes the relationship between μ and θ
- ▶ A sample-size-indifferent construction asks for the prior predictive distribution of t not to depend on n

Seek a prior for μ that expresses indifference to sample size in this problem

THE OBJECTIVE-PRIOR IDEA

The sample-size-indifference requirement leads to a scale-invariant prior of the form

$$\pi(\mu) \propto \frac{1}{|\mu|}.$$

- ▶ Formally, this construction removes the sample-size dependence from the comparison
- ▶ This prior is **improper**, however — it cannot be normalized as a probability distribution
- ▶ Consequently, it cannot be used naively to obtain a well-defined marginal likelihood and Bayes factor
- ▶ The apparent removal of sample-size dependence comes with a normalization problem

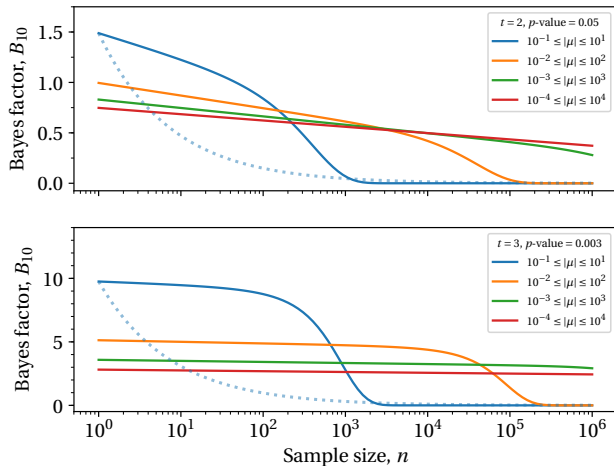
TRUNCATING THE PRIOR

A practical alternative is to truncate the scale-invariant prior, for example by restricting the range of plausible values of $|\mu|$

- ▶ Truncation makes the prior proper
- ▶ The resulting evidence again depends on sample size
- ▶ This dependence can be delayed, potentially making the effect practically relevant

Objective sample-size indifference is not obtained for free: proper truncation restores some dependence on n

THE JEFFREYS-LINDLEY PARADOX



- ▶ Increase n while keeping the test statistic t fixed
- ▶ The p -value remains unchanged
- ▶ Bayes factor starts constant — the paradox delayed by scale-invariant prior
- ▶ Because of the truncation, increasing n eventually leads to strong Bayesian evidence for the null

WHAT DOES THE PARADOX TELL US?

- ▶ The paradox is not a mathematical contradiction
- ▶ The p -value and Bayes factor summarize evidence differently
- ▶ Prior assumptions matter, especially for model comparison
- ▶ The scale-invariant construction links the paradox to the impropriety of the prior

Invariance can clarify a modelling choice, but it does not remove the need to examine prior propriety and the meaning of the resulting evidence

WHAT IS THE QUESTION?

If the Bayesian answer was troubling, we should carefully think about what question we were asking (Jaynes, 1980)

- ▶ At first, we might think we were asking whether $\mu = 0$ or $\mu \neq 0$ — this is classical hypothesis test formulation
- ▶ A proper prior, however, cannot be scale invariant; at least, it must disfavour arbitrarily tiny and enormous scales
- ▶ Thus, we are really asking **is $\mu = 0$ or is it at the scales favoured by the choice of prior?**

Asked this way, it is intuitive that for fixed t the preference for $\mu = 0$ must eventually grow as n increases, as eventually the data favour μ smaller than that favoured by the prior

SECTION 4

SUMMARY

SUMMARY

- ▶ Paradoxes shed light on the foundations and assumptions in a topic
- ▶ The Jeffreys–Lindley paradox contrasts fixed- p -value reasoning with Bayesian model comparison as n grows
- ▶ Requiring sample-size indifference leads to a scale-invariant, improper prior
- ▶ A truncated scale-invariant prior can delay, but does not eliminate, sample-size dependence
- ▶ The prior and the inferential question are central to interpreting the result
- ▶ In the Jeffreys–Lindley paradox, we are really asking **is $\mu = 0$ or is it at the scales favoured by the choice of prior?**

See Fowlie A., (2020). *Objective Bayesian approach to the Jeffreys-Lindley paradox*, *Communications in Statistics – Theory and Methods* **51** 6760–6765

REFERENCES I

Bertrand J., (1888). *Calcul des probabilités*. Gauthier-Villars Et Fils, Paris.

Fowlie A., (2020). *Objective Bayesian approach to the Jeffreys-Lindley paradox*,
Communications in Statistics – Theory and Methods **51** 6760–6765 [2012.04879].

Jaynes E. T., (1973). *The well-posed problem*, *Foundations of Physics* **3** 477–492.

Jaynes E. T., (1980). *A Bayesian Analysis of Classical Hypothesis Testing by José M. Bernardo—Discussion*, *Trabajos de Estadística Y de Investigación Operativa* **31** 605–647.

Jeffreys H., (1939). *Theory of Probability*. Oxford University Press.

Lindley D. V., (1957). *A statistical paradox*, *Biometrika* **44** 187–192.

REFERENCES II

Selvin S., (1975). *A problem in probability (letter to the editor)*, *The American Statistician* **29** 67.

Simpson E. H., (1951). *The interpretation of interaction in contingency tables*, *Journal of the Royal Statistical Society: Series B* **13** 238.