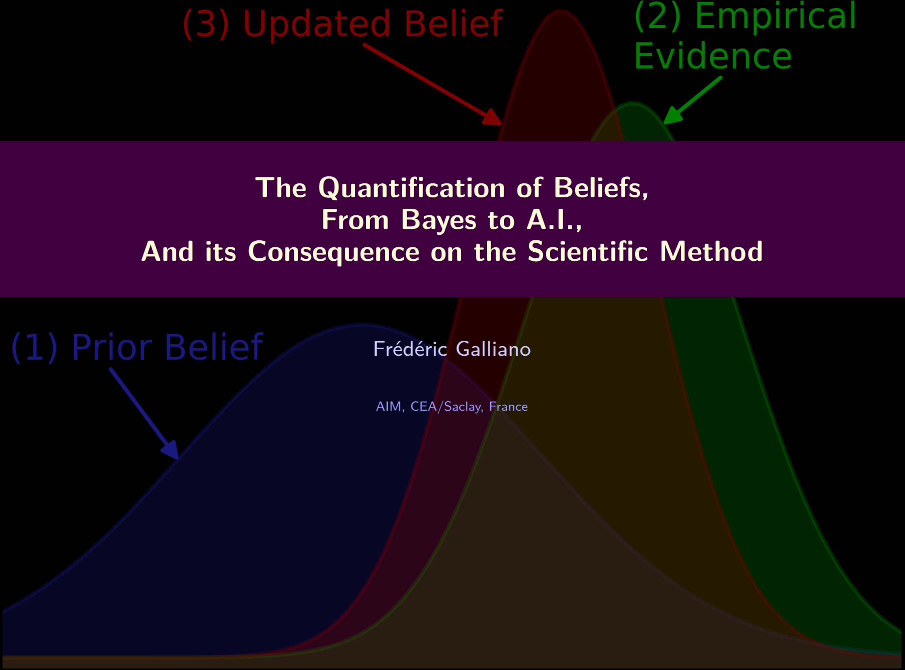


The Quantification of Beliefs,
From Bayes to A.I.,
And its Consequence on the Scientific Method

Frédéric Galliano

AIM, CEA/Saclay, France

Outline of the Talk

1. **BAYESIANS VS FREQUENTISTS**
   - Epistemological Principles & Comparison
   - Demonstration on a Simple Example
   - Limitations of the Frequentist Approach

Outline of the Talk

1. **BAYESIANS VS FREQUENTISTS**
   - Epistemological Principles & Comparison
   - Demonstration on a Simple Example
   - Limitations of the Frequentist Approach

2. **BAYES' RULE THROUGH HISTORY**
   - Early Development
   - The Frequentist Winter
   - The Bayesian Renaissance

Outline of the Talk

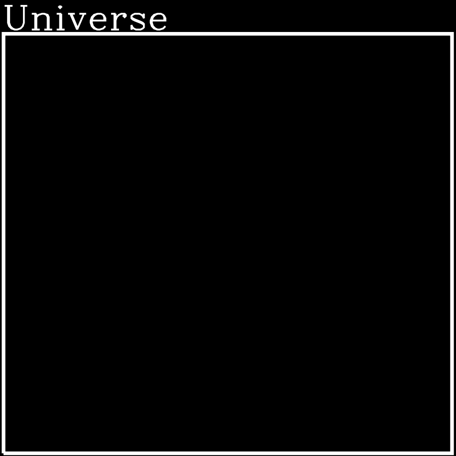
1. **BAYESIANS VS FREQUENTISTS**
   - Epistemological Principles & Comparison
   - Demonstration on a Simple Example
   - Limitations of the Frequentist Approach
2. **BAYES' RULE THROUGH HISTORY**
   - Early Development
   - The Frequentist Winter
   - The Bayesian Renaissance
3. **IMPLICATIONS FOR THE SCIENTIFIC METHOD**
   - Karl Popper's Logic of Scientific Discovery
   - Bayesian Epistemology
   - How Researchers Actually Work

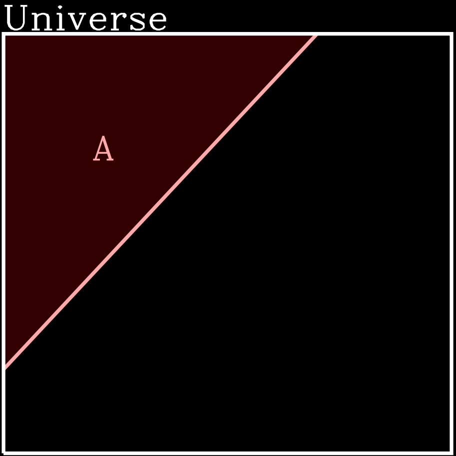
Outline of the Talk

Outline of the Talk

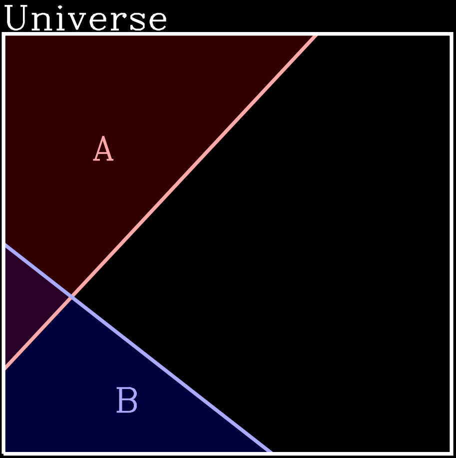
Prologue: Visual Representation of Bayes' Rule (Venn Diagram)

Prologue: Visual Representation of Bayes' Rule (Venn Diagram)

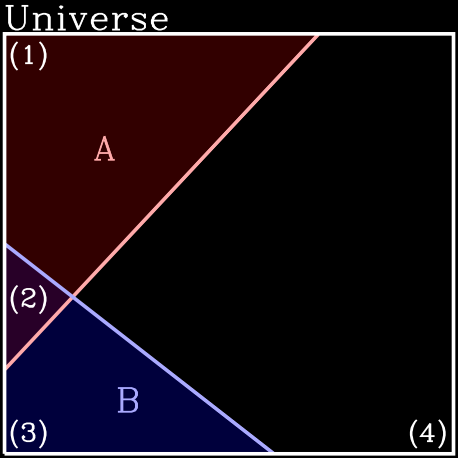


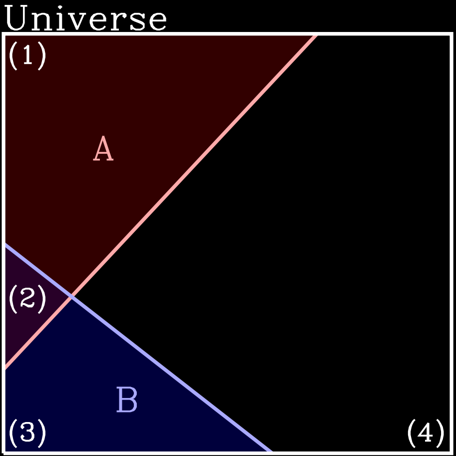

Universe
A

Prologue: Visual Representation of Bayes' Rule (Venn Diagram)



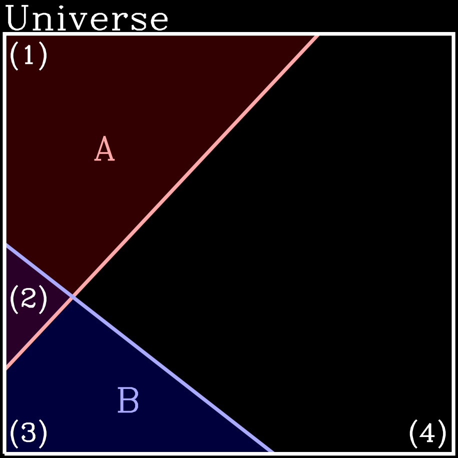
Prologue: Visual Representation of Bayes' Rule (Venn Diagram)





Conditional probability:

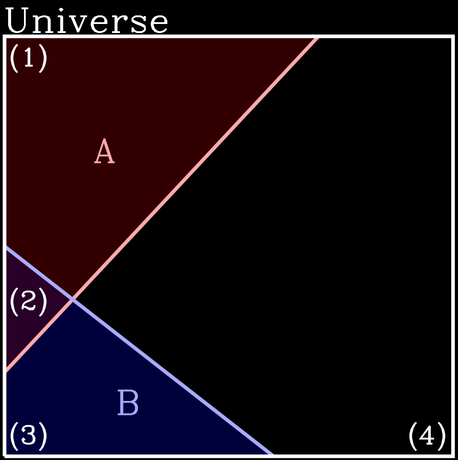
$p(A|B)$: probability of A, knowing B



Conditional probability:

$p(A|B)$: probability of A, knowing B
$\Leftrightarrow$ probability of A, within B (new Universe):

Prologue: Visual Representation of Bayes' Rule (Venn Diagram)

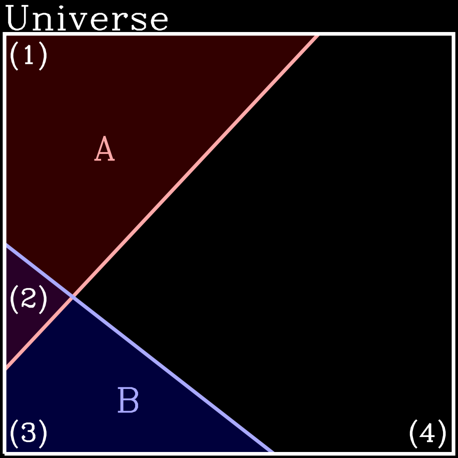


Conditional probability:

$p(A|B)$: probability of A, knowing B
$\Leftrightarrow$ probability of A, within B (new Universe):

$$p(A|B) = \frac{(2)}{(2)+(3)}$$

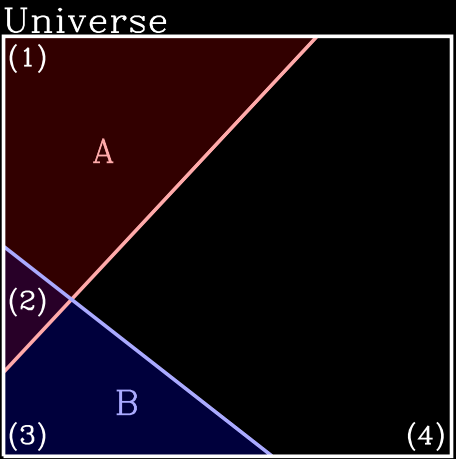
Prologue: Visual Representation of Bayes' Rule (Venn Diagram)



Conditional probability:

$p(A|B)$: probability of A, knowing B
$\Leftrightarrow$ probability of A, within B (new Universe):

$$p(A|B) = \frac{(2)}{(2)+(3)} = \frac{p(A \text{ and } B)}{p(B)}.$$



Conditional probability:

$p(A|B)$: probability of A, knowing B
$\Leftrightarrow$ probability of A, within B (new Universe):

$$p(A|B) = \frac{(2)}{(2)+(3)} = \frac{p(A \text{ and } B)}{p(B)}.$$

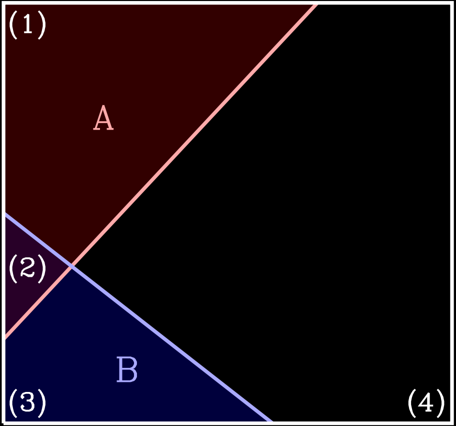
Bayes' rule (general probability theorem):

By symmetry of A and B:

$$p(A \text{ and } B) = p(A|B)p(B)$$

Prologue: Visual Representation of Bayes' Rule (Venn Diagram)

Universe



Conditional probability:

$p(A|B)$: probability of A, knowing B
$\Leftrightarrow$ probability of A, within B (new Universe):

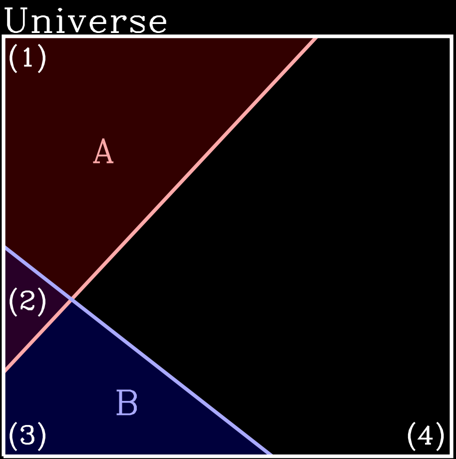
$$p(A|B) = \frac{(2)}{(2)+(3)} = \frac{p(A \text{ and } B)}{p(B)}.$$

Bayes' rule (general probability theorem):

By symmetry of A and B:

$$p(A \text{ and } B) = p(A|B)p(B) = p(B|A)p(A)$$

Prologue: Visual Representation of Bayes' Rule (Venn Diagram)



Conditional probability:

$p(A|B)$: probability of A, knowing B

$\Leftrightarrow$ probability of A, within B (new Universe):

$$p(A|B) = \frac{(2)}{(2)+(3)} = \frac{p(A \text{ and } B)}{p(B)}.$$

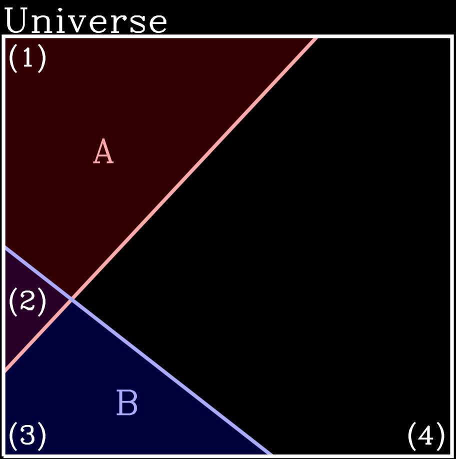
Bayes' rule (general probability theorem):

By symmetry of A and B:

$$p(A \text{ and } B) = p(A|B)p(B) = p(B|A)p(A)$$

$$\Rightarrow p(A|B) = \frac{p(B|A)p(A)}{p(B)}.$$

Prologue: Visual Representation of Bayes' Rule (Venn Diagram)



Conditional probability:

$p(A|B)$: probability of A, knowing B
$\Leftrightarrow$ probability of A, within B (new Universe):

$$p(A|B) = \frac{(2)}{(2)+(3)} = \frac{p(A \text{ and } B)}{p(B)}.$$

Bayes' rule (general probability theorem):

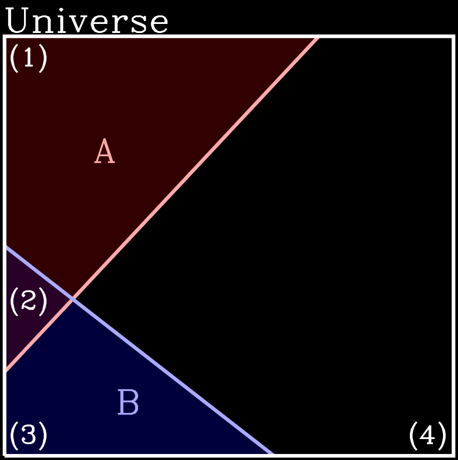
By symmetry of A and B:

$$p(A \text{ and } B) = p(A|B)p(B) = p(B|A)p(A)$$

$$\Rightarrow p(A|B) = \frac{p(B|A)p(A)}{p(B)}.$$

Example of $p(A|B) \neq p(B|A)$:

Prologue: Visual Representation of Bayes' Rule (Venn Diagram)



Conditional probability:

$p(A|B)$: probability of A, knowing B
$\Leftrightarrow$ probability of A, within B (new Universe):

$$p(A|B) = \frac{(2)}{(2)+(3)} = \frac{p(A \text{ and } B)}{p(B)}.$$

Bayes' rule (general probability theorem):

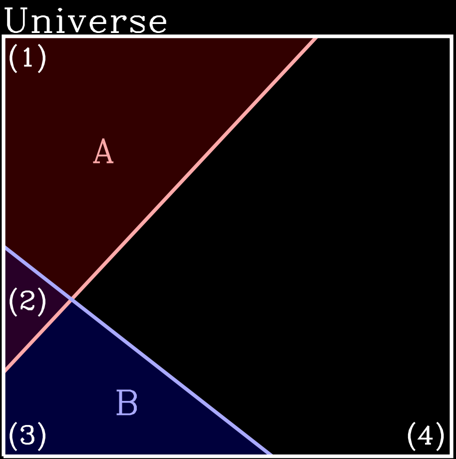
By symmetry of A and B:

$$p(A \text{ and } B) = p(A|B)p(B) = p(B|A)p(A)$$

$$\Rightarrow p(A|B) = \frac{p(B|A)p(A)}{p(B)}.$$

Example of $p(A|B) \neq p(B|A)$:

$$p(\text{SNIa}|\text{binary}) \simeq 10^{-12}\ \text{yr}^{-1}$$



Conditional probability:

$p(A|B)$: probability of A, knowing B
$\Leftrightarrow$ probability of A, within B (new Universe):

$$p(A|B) = \frac{(2)}{(2)+(3)} = \frac{p(A \text{ and } B)}{p(B)}.$$

Bayes' rule (general probability theorem):

By symmetry of A and B:

$$p(A \text{ and } B) = p(A|B)p(B) = p(B|A)p(A)$$

$$\Rightarrow p(A|B) = \frac{p(B|A)p(A)}{p(B)}.$$

Example of $p(A|B) \neq p(B|A)$:

$$p(\text{SNIa}|\text{binary}) \simeq 10^{-12}\ \mathrm{yr}^{-1} \quad \text{while} \quad p(\text{binary}|\text{SNIa}) = 1.$$

Two Conceptions of Probability & Uncertainty

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

$\Rightarrow$ probabilities *can* be assigned to parameters and hypotheses.

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

$\Rightarrow$ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

$\Rightarrow$ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}}$$

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

$\Rightarrow$ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}} \propto \underbrace{p(data|par)}_{\text{likelihood}} \times$$

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

$\Rightarrow$ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}} \propto \underbrace{p(data|par)}_{\text{likelihood}} \times \underbrace{p(par)}_{\text{prior}}.$$

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

$\Rightarrow$ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}} \propto \underbrace{p(data|par)}_{\text{likelihood}} \times \underbrace{p(par)}_{\text{prior}}.$$

Results: the posterior contains all the *relevant information*.

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

$\Rightarrow$ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}} \propto \underbrace{p(data|par)}_{\text{likelihood}} \times \underbrace{p(par)}_{\text{prior}}.$$

Results: the posterior contains all the *relevant information*.

$\Rightarrow$ use it to give credible ranges, test hypotheses, compare models, *etc.*

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

$\Rightarrow$ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}} \propto \underbrace{p(data|par)}_{\text{likelihood}} \times \underbrace{p(par)}_{\text{prior}}.$$

Results: the posterior contains all the *relevant information*.

$\Rightarrow$ use it to give credible ranges, test hypotheses, compare models, *etc.*

The Frequentist View

There is an objective truth, and science should not deal with subjective notions.

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

⇒ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}} \propto \underbrace{p(data|par)}_{\text{likelihood}} \times \underbrace{p(par)}_{\text{prior}}.$$

Results: the posterior contains all the *relevant information*.

⇒ use it to give credible ranges, test hypotheses, compare models, *etc.*

The Frequentist View

There is an objective truth, and science should not deal with subjective notions.

Probability: limit to infinity of the *occurence frequency* of an event, in a sequence of repeated experiments.

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

$\Rightarrow$ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}} \propto \underbrace{p(data|par)}_{\text{likelihood}} \times \underbrace{p(par)}_{\text{prior}}.$$

Results: the posterior contains all the *relevant information*.

$\Rightarrow$ use it to give credible ranges, test hypotheses, compare models, *etc.*

The Frequentist View

There is an objective truth, and science should not deal with subjective notions.

Probability: limit to infinity of the *occurence frequency* of an event, in a sequence of repeated experiments.

$\Rightarrow$ probabilities *can not* be assigned to parameters or hypotheses.

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

⇒ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}} \propto \underbrace{p(data|par)}_{\text{likelihood}} \times \underbrace{p(par)}_{\text{prior}}.$$

Results: the posterior contains all the *relevant information*.

⇒ use it to give credible ranges, test hypotheses, compare models, *etc.*

The Frequentist View

There is an objective truth, and science should not deal with subjective notions.

Probability: limit to infinity of the *occurence frequency* of an event, in a sequence of repeated experiments.

⇒ probabilities *can not* be assigned to parameters or hypotheses.

Inference: sample the likelihood, conditional on the parameters:

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

$\Rightarrow$ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}} \propto \underbrace{p(data|par)}_{\text{likelihood}} \times \underbrace{p(par)}_{\text{prior}}.$$

Results: the posterior contains all the *relevant information*.

$\Rightarrow$ use it to give credible ranges, test hypotheses, compare models, *etc.*

The Frequentist View

There is an objective truth, and science should not deal with subjective notions.

Probability: limit to infinity of the *occurence frequency* of an event, in a sequence of repeated experiments.

$\Rightarrow$ probabilities *can not* be assigned to parameters or hypotheses.

Inference: sample the likelihood, conditional on the parameters:

$$L(par) \equiv p(data|par).$$

Two Conceptions of Probability & Uncertainty

The Bayesian View

There is an objective truth, but our *knowledge* is only *partial and subjective*.

Probability: quantification of the *plausibility* of a proposition, with incomplete knowledge (subjective point of view).

⇒ probabilities *can* be assigned to parameters and hypotheses.

Inference: sample the distribution of parameters, conditional on the data (Bayes' rule):

$$\underbrace{p(par|data)}_{\text{posterior}} \propto \underbrace{p(data|par)}_{\text{likelihood}} \times \underbrace{p(par)}_{\text{prior}}.$$

Results: the posterior contains all the *relevant information*.

⇒ use it to give credible ranges, test hypotheses, compare models, *etc.*

The Frequentist View

There is an objective truth, and science should not deal with subjective notions.

Probability: limit to infinity of the *occurence frequency* of an event, in a sequence of repeated experiments.

⇒ probabilities *can not* be assigned to parameters or hypotheses.

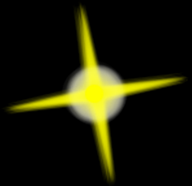
Inference: sample the likelihood, conditional on the parameters:

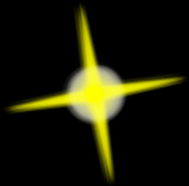
$$L(par) \equiv p(data|par).$$

Results: describe how the derived parameter value would vary if we were to *repeat the experiment* in the same conditions.

Comparison of the Two Approaches on a Simple Case

Comparison of the Two Approaches on a Simple Case

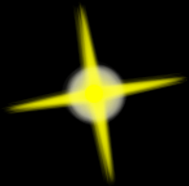




Observations

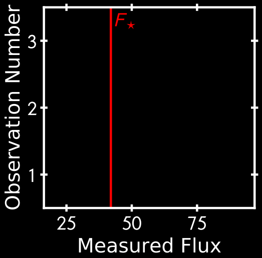
- True flux, $F_\star = 42$;

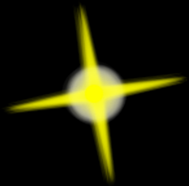
Comparison of the Two Approaches on a Simple Case



Observations

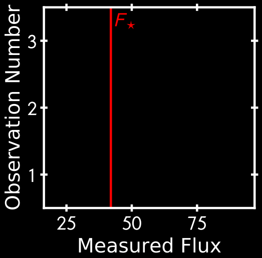
- True flux, $F_\star = 42$;



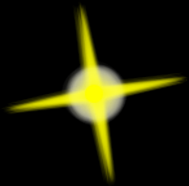


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;

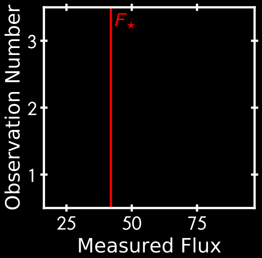


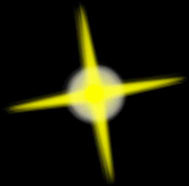
Comparison of the Two Approaches on a Simple Case



Observations

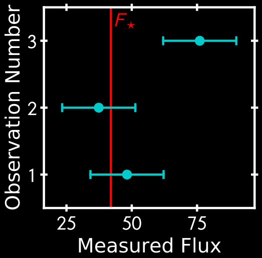
- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.



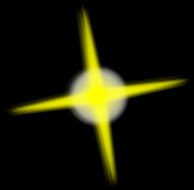


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.



Comparison of the Two Approaches on a Simple Case

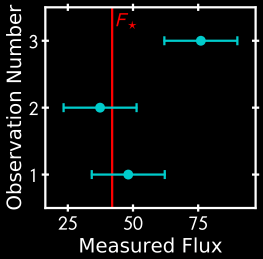


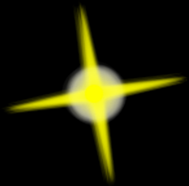
Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}}$$



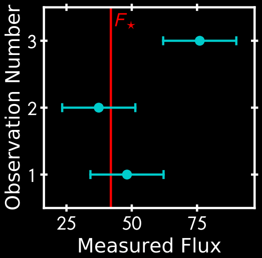


Observations

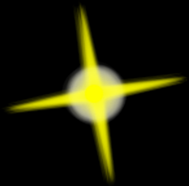
- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^N F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Comparison of the Two Approaches on a Simple Case

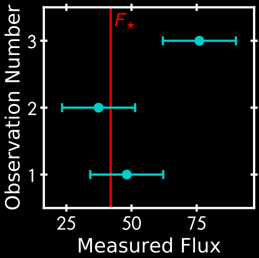


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

$\Rightarrow$ classic solution:

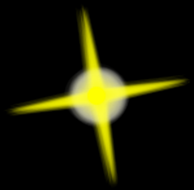
$$F_\star \simeq \frac{\sum_{i=1}^N F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times$$

Comparison of the Two Approaches on a Simple Case

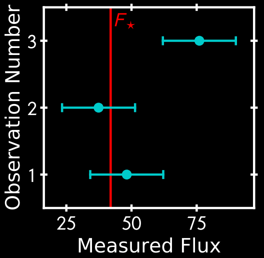


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

$\Rightarrow$ classic solution:

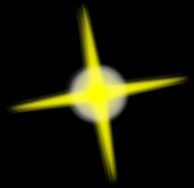
$$F_\star \simeq \frac{\sum_{i=1}^N F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

Comparison of the Two Approaches on a Simple Case

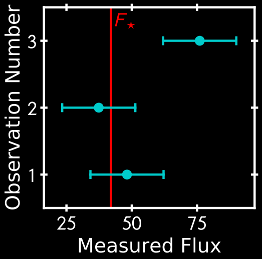


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

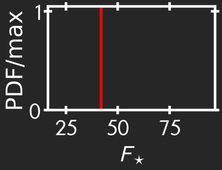
$\Rightarrow$ classic solution:

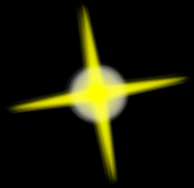
$$F_\star \simeq \frac{\sum_{i=1}^N F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$



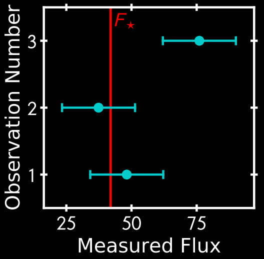


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

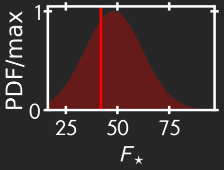
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$

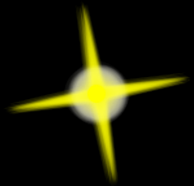


Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$



Comparison of the Two Approaches on a Simple Case

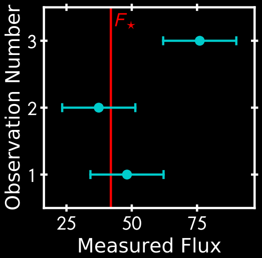


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

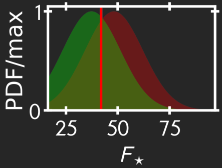
$\Rightarrow$ classic solution:

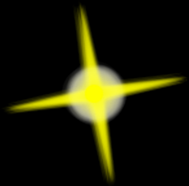
$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$



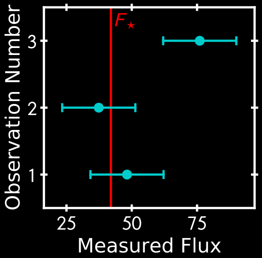


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

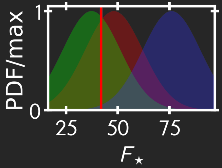
$\Rightarrow$ classic solution:

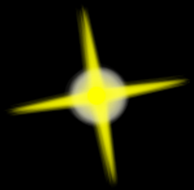
$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$



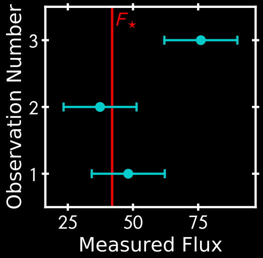


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

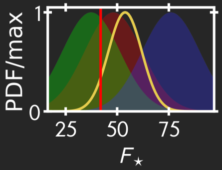
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$

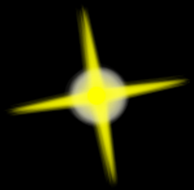


Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$



Comparison of the Two Approaches on a Simple Case

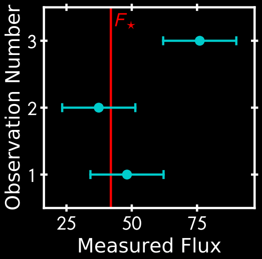


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

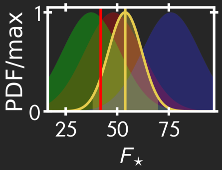
$\Rightarrow$ classic solution:

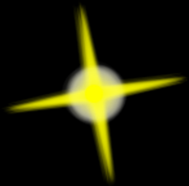
$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star | F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$



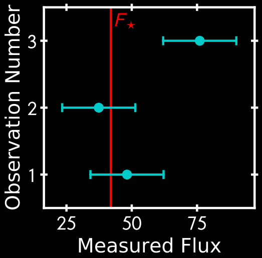


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

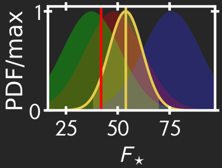
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

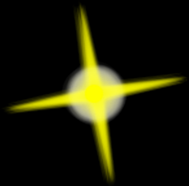
$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$



MCMC:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Comparison of the Two Approaches on a Simple Case

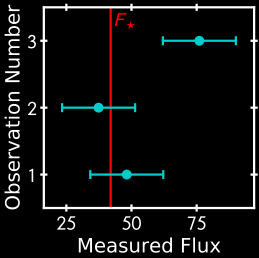


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

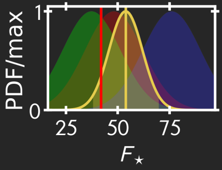
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$



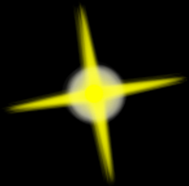
MCMC:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, $F_{\rm ML}$;

Comparison of the Two Approaches on a Simple Case

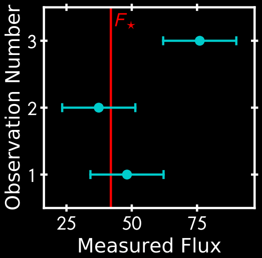


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

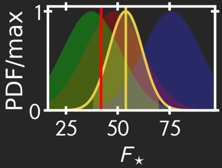
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^N F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$



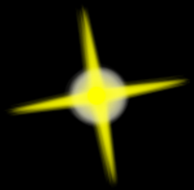
MCMC:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, F_{ML};
2. Sampling $F_i \Rightarrow$ confidence interval.

Comparison of the Two Approaches on a Simple Case

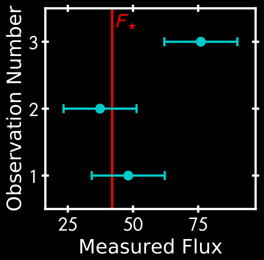


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

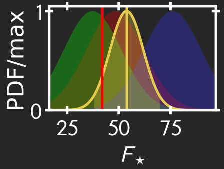
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star | F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

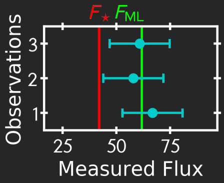


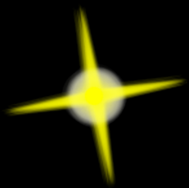
MCMC:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, $F_{\rm ML}$;
2. Sampling $F_i \Rightarrow$ confidence interval.



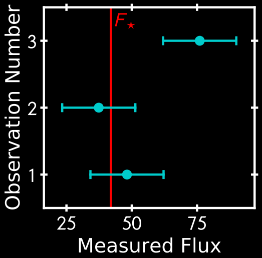


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

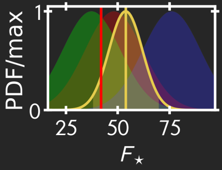
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

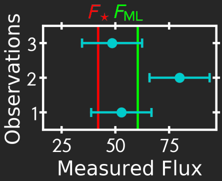


MCMC:

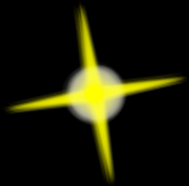
- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, F_{ML};
2. Sampling $F_i \Rightarrow$ confidence interval.



Comparison of the Two Approaches on a Simple Case

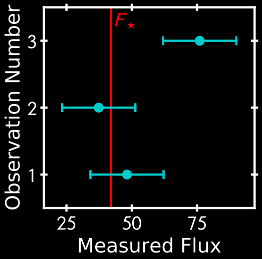


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

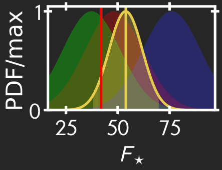
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

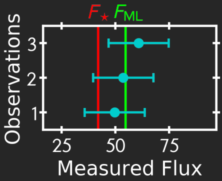


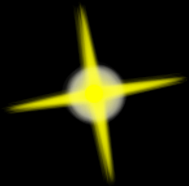
MCMC:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, $F_{\rm ML}$;
2. Sampling $F_i \Rightarrow$ confidence interval.



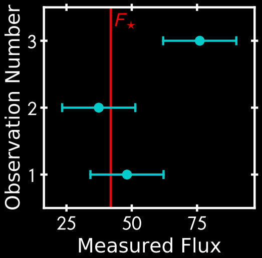


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

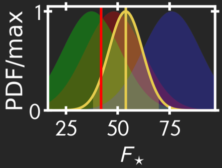
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

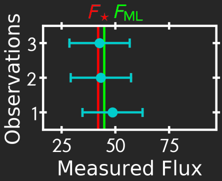


MCMC:

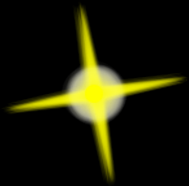
- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, $F_{\rm ML}$;
2. Sampling $F_i \Rightarrow$ confidence interval.



Comparison of the Two Approaches on a Simple Case

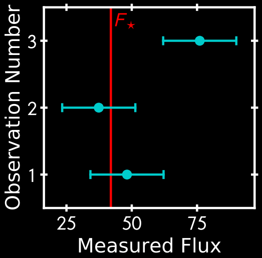


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

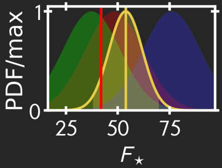
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

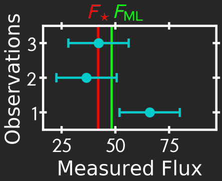


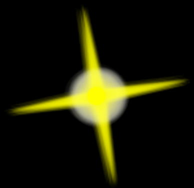
MCMC:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, F_{ML};
2. Sampling $F_i \Rightarrow$ confidence interval.



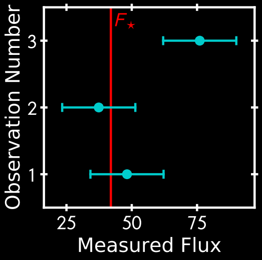


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

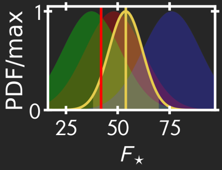
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^N F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

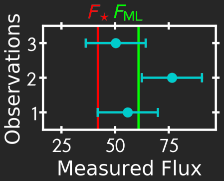


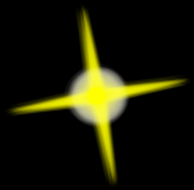
MCMC:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, F_{ML};
2. Sampling $F_i \Rightarrow$ confidence interval.



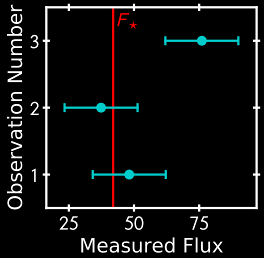


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

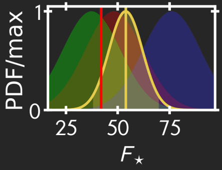
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^N F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

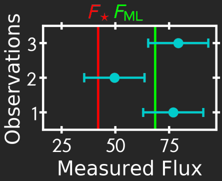


MCMC:

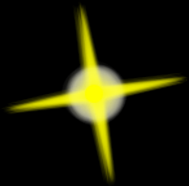
- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, $F_{\rm ML}$;
2. Sampling $F_i \Rightarrow$ confidence interval.



Comparison of the Two Approaches on a Simple Case

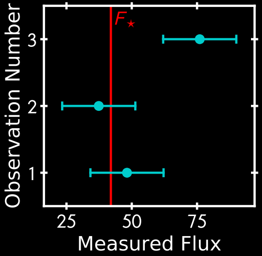


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

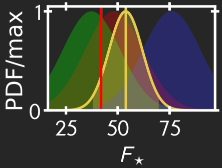
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

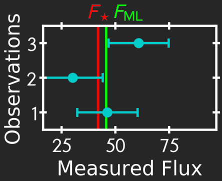


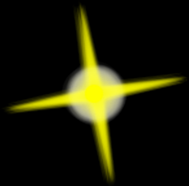
MCMC:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, F_{ML};
2. Sampling $F_i \Rightarrow$ confidence interval.



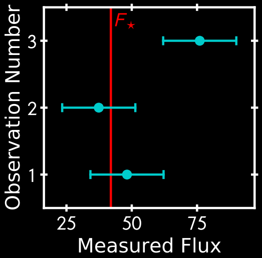


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

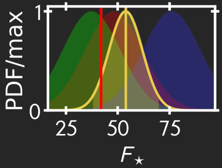
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

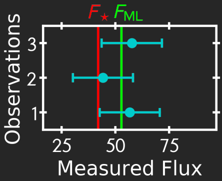


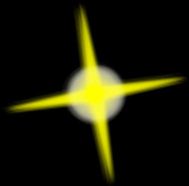
MCMC:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, F_{ML};
2. Sampling $F_i \Rightarrow$ confidence interval.



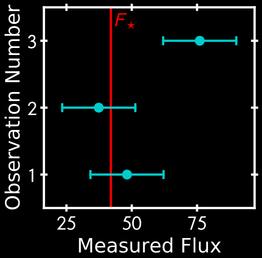


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

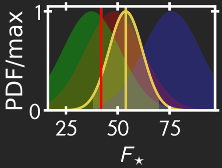
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^{N} F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

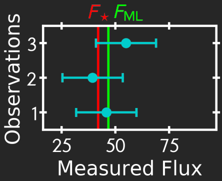


MCMC:

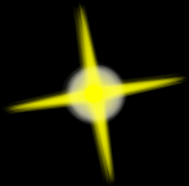
- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, $F_{\rm ML}$;
2. Sampling $F_i \Rightarrow$ confidence interval.



Comparison of the Two Approaches on a Simple Case

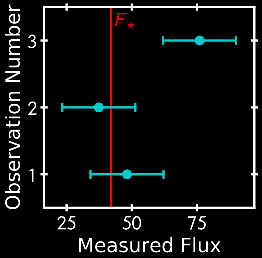


Observations

- True flux, $F_\star = 42$;
- $N = 3$ observations, F_i;
- Gaussian noise, $\sigma_F = 14$.

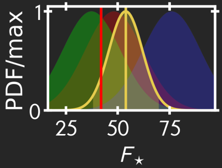
$\Rightarrow$ classic solution:

$$F_\star \simeq \frac{\sum_{i=1}^N F_i}{N} \pm \frac{\sigma_F}{\sqrt{N}} = 53.9 \pm 8.1.$$



Bayesian Solution:

$$p(F_\star|F_i) \propto \prod_i \exp\left(-\frac{(F_\star - F_i)^2}{2\sigma_F^2}\right) \times \underbrace{1}_{\text{prior}}$$

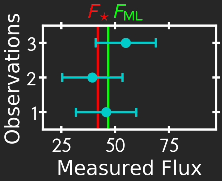


MCMC:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % credible range: $[37.8, 69.8]$.

Frequentist Solution:

1. Maximum-likelihood, $F_{\rm ML}$;
2. Sampling $F_i \Rightarrow$ confidence interval.



Bootstrapping:

- $F_\star \simeq 53.9 \pm 8.1$;
- 95 % confidence interval: $[37.8, 69.8]$.

Difference Between the Two Approaches: Using an Informative Prior

Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

Difference Between the Two Approaches: Using an Informative Prior

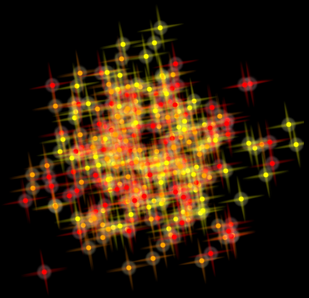
Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

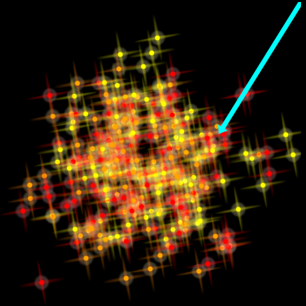
Example: the star is in a globular cluster & you know the distance.



Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

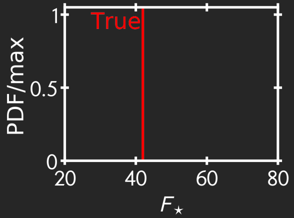


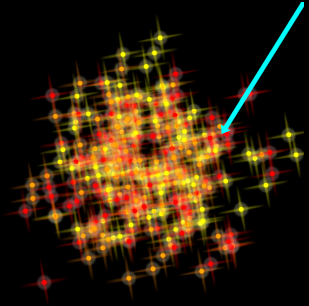
Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

Prior: cluster mass function & mass/luminosity relation ⇒ luminosity function.



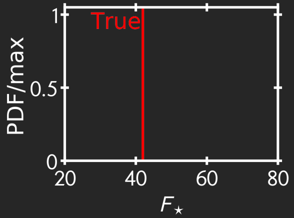


Difference Between the Two Approaches: Using an Informative Prior

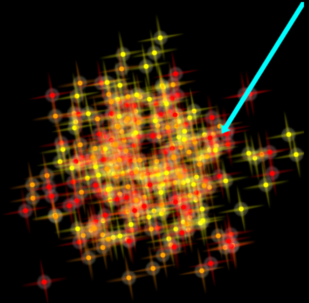
Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

Prior: cluster mass function & mass/luminosity relation ⇒ luminosity function.



True flux: $F_\star = 42$

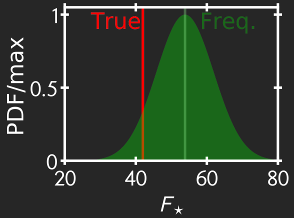


Difference Between the Two Approaches: Using an Informative Prior

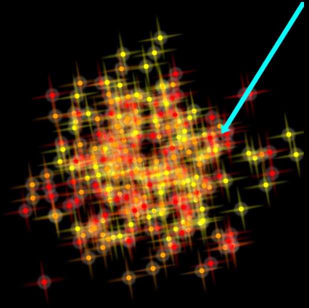
Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

Prior: cluster mass function & mass/luminosity relation $\Rightarrow$ luminosity function.



True flux: $F_\star = 42$

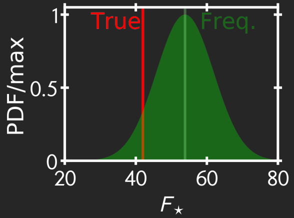


Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

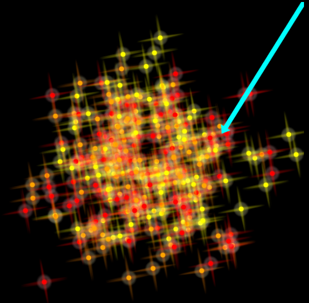
Example: the star is in a globular cluster & you know the distance.

Prior: cluster mass function & mass/luminosity relation $\Rightarrow$ luminosity function.



True flux: $F_\star = 42$

Frequentist: $F_\star \simeq 53.9 \pm 8.1$

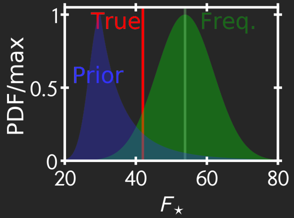


Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

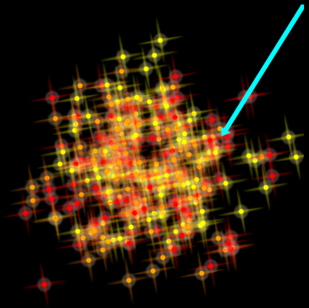
Example: the star is in a globular cluster & you know the distance.

Prior: cluster mass function & mass/luminosity relation ⇒ luminosity function.



True flux: $F_\star = 42$

Frequentist: $F_\star \simeq 53.9 \pm 8.1$

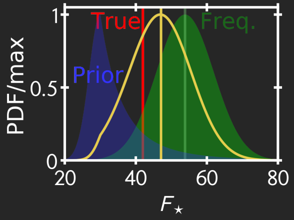


Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

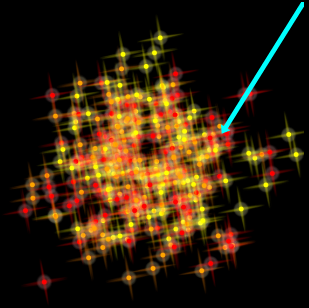
Example: the star is in a globular cluster & you know the distance.

Prior: cluster mass function & mass/luminosity relation ⇒ luminosity function.



True flux: $F_\star = 42$

Frequentist: $F_\star \simeq 53.9 \pm 8.1$

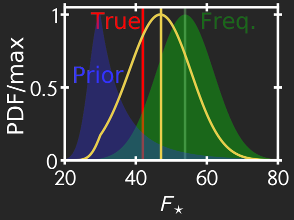


Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

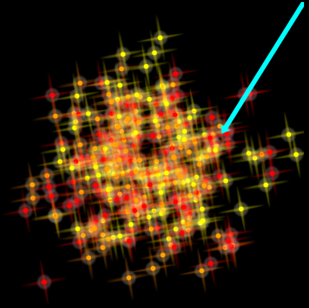
Prior: cluster mass function & mass/luminosity relation ⇒ luminosity function.



True flux: $F_\star = 42$

Frequentist: $F_\star \simeq 53.9 \pm 8.1$

Bayesian: $F_\star \simeq 47.1 \pm 8.0$

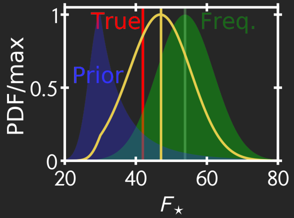


Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

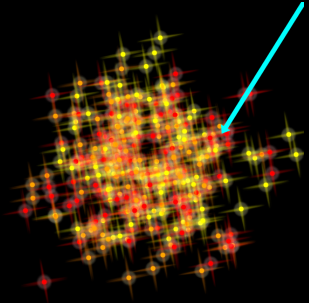
Prior: cluster mass function & mass/luminosity relation ⇒ luminosity function.



True flux: $F_\star = 42$

Frequentist: $F_\star \simeq 53.9 \pm 8.1$

Bayesian: $F_\star \simeq 47.1 \pm 8.0$



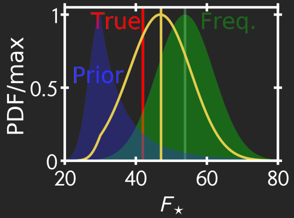
Refining the Prior Based on the Observations:

Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

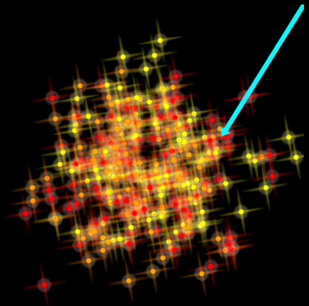
Prior: cluster mass function & mass/luminosity relation $\Rightarrow$ luminosity function.



True flux: $F_\star = 42$

Frequentist: $F_\star \simeq 53.9 \pm 8.1$

Bayesian: $F_\star \simeq 47.1 \pm 8.0$



Refining the Prior Based on the Observations:

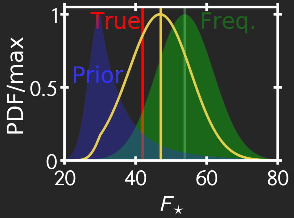
Cumulation of data: if you perform a series of observations in this cluster:

Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

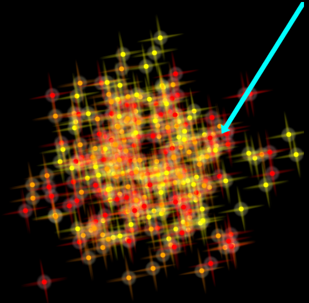
Prior: cluster mass function & mass/luminosity relation ⇒ luminosity function.



True flux: $F_\star = 42$

Frequentist: $F_\star \simeq 53.9 \pm 8.1$

Bayesian: $F_\star \simeq 47.1 \pm 8.0$



Refining the Prior Based on the Observations:

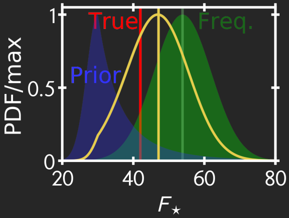
Cumulation of data: if you perform a series of observations in this cluster:

$$\underbrace{p(F_\star)}_{\text{new prior}}$$

Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

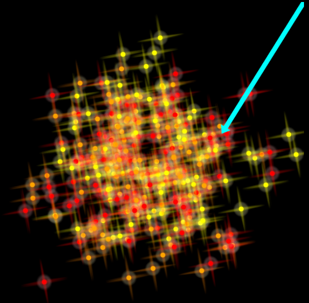
Prior: cluster mass function & mass/luminosity relation ⇒ luminosity function.



True flux: $F_\star = 42$

Frequentist: $F_\star \simeq 53.9 \pm 8.1$

Bayesian: $F_\star \simeq 47.1 \pm 8.0$



Refining the Prior Based on the Observations:

Cumulation of data: if you perform a series of observations in this cluster:

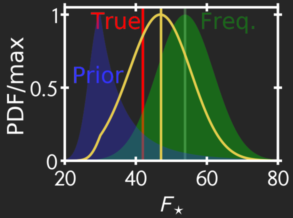
$$\underbrace{p(F_\star)}_{\text{new prior}} \propto \underbrace{1}_{\text{initial prior}} \times$$

Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

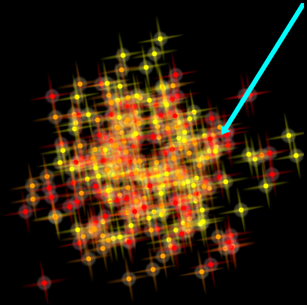
Prior: cluster mass function & mass/luminosity relation ⇒ luminosity function.



True flux: $F_\star = 42$

Frequentist: $F_\star \simeq 53.9 \pm 8.1$

Bayesian: $F_\star \simeq 47.1 \pm 8.0$



Refining the Prior Based on the Observations:

Cumulation of data: if you perform a series of observations in this cluster:

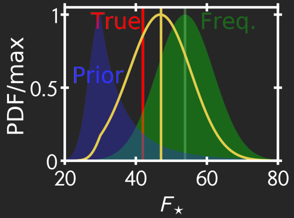
$$\underbrace{p(F_\star)}_{\text{new prior}} \propto \underbrace{1}_{\text{initial prior}} \times \underbrace{p(F_\star^{(1)}|F_\star) \times \ldots \times p(F_\star^{(N_\star)}|F_\star)}_{\text{cumulated data}}$$

Difference Between the Two Approaches: Using an Informative Prior

Accounting for Additional Information:

Example: the star is in a globular cluster & you know the distance.

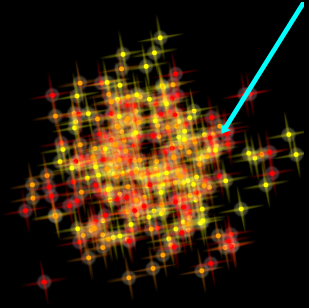
Prior: cluster mass function & mass/luminosity relation ⇒ luminosity function.



True flux: $F_\star = 42$

Frequentist: $F_\star \simeq 53.9 \pm 8.1$

Bayesian: $F_\star \simeq 47.1 \pm 8.0$



Refining the Prior Based on the Observations:

Cumulation of data: if you perform a series of observations in this cluster:

$$\underbrace{p(F_\star)}_{\text{new prior}} \propto \underbrace{1}_{\text{initial prior}} \times \underbrace{p(F_\star^{(1)}|F_\star) \times \ldots \times p(F_\star^{(N_\star)}|F_\star)}_{\text{cumulated data}}$$

Hierarchical model: consistently perform this process on all the data, at the same time.

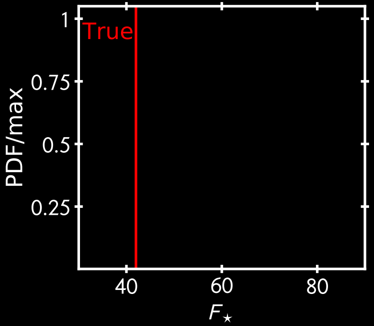
Difference Between the Two Approaches: Lindley's Paradox

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

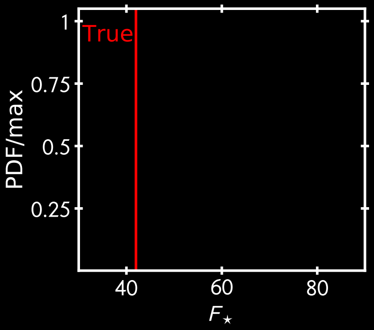


(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

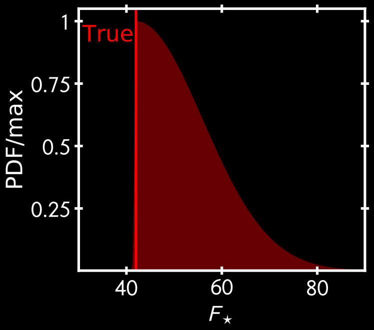
Noise: heavily-skewed split-normal distribution.



(Example adapted from Jaynes, 1976)

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

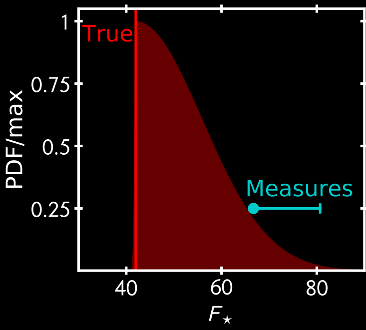


(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

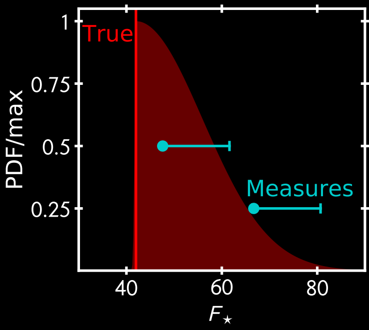


(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

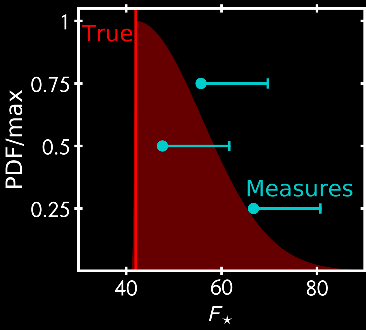


(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

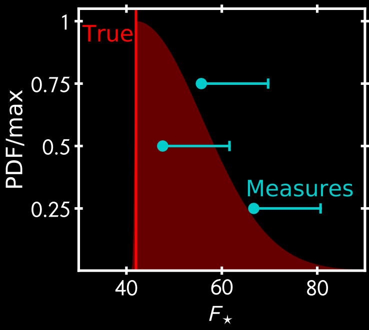


(Example adapted from Jaynes, 1976)

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

Common sense: true flux < minimum measure: $F_\star \lesssim 47.6$.



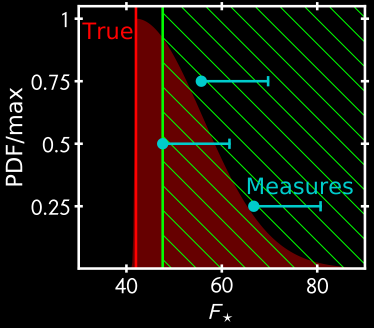
(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

Common sense: true flux $<$ minimum measure: $F_\star \lesssim 47.6$.



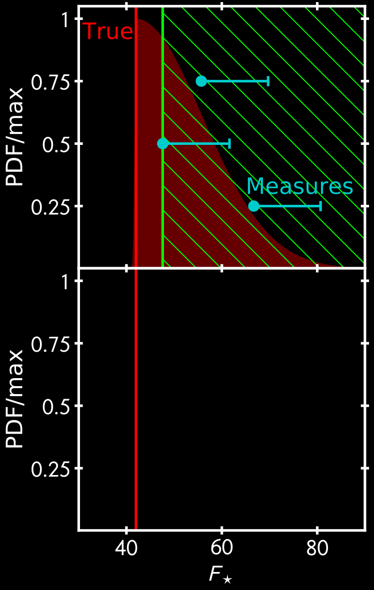
(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

Common sense: true flux < minimum measure: $F_\star \lesssim 47.6$.



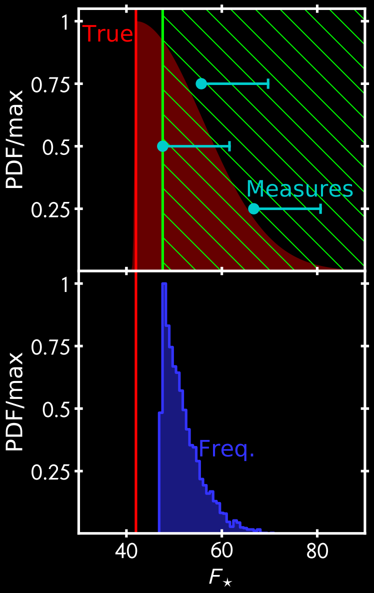
(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

Common sense: true flux < minimum measure: $F_\star \lesssim 47.6$.



(Example adapted from Jaynes, 1976)

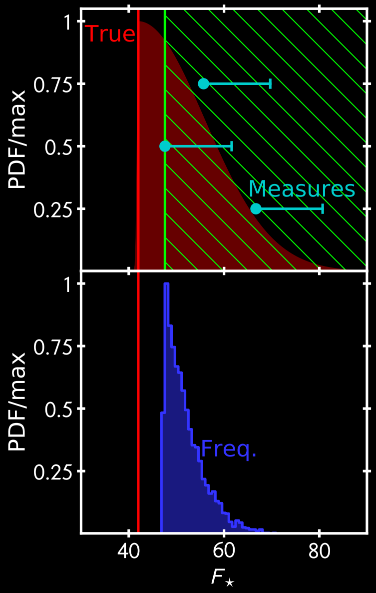
Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

Common sense: true flux $<$ minimum measure: $F_\star \lesssim 47.6$.

Frequentism: 95 % confidence interval: $[47.5, 62.5]$ $\Rightarrow$ inconsistent solution.



(Example adapted from Jaynes, 1976)

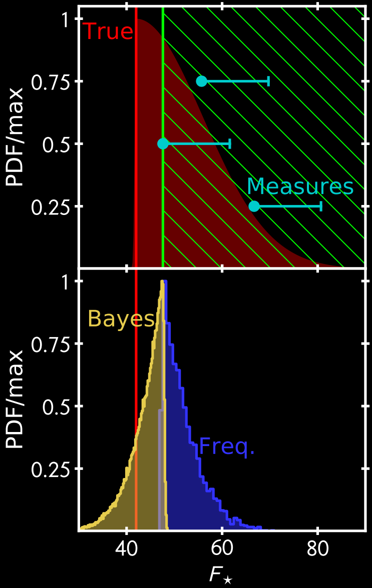
Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

Common sense: true flux < minimum measure: $F_\star \lesssim 47.6$.

Frequentism: 95 % confidence interval: $[47.5, 62.5]$
$\Rightarrow$ inconsistent solution.



(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

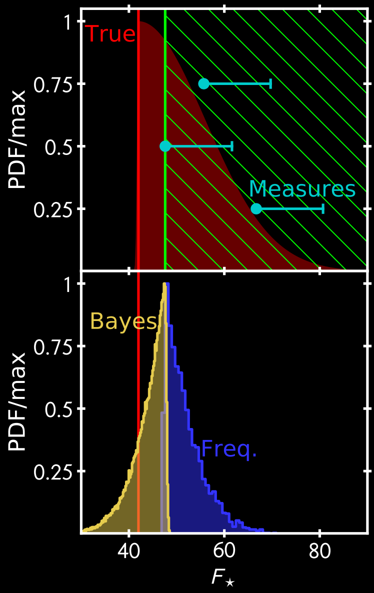
Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

Common sense: true flux < minimum measure: $F_\star \lesssim 47.6$.

Frequentism: 95 % confidence interval: [47.5, 62.5] $\Rightarrow$ inconsistent solution.

Bayesianism: 95 % credible range: [34.6, 47.7] $\Rightarrow$ consistent solution.



(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

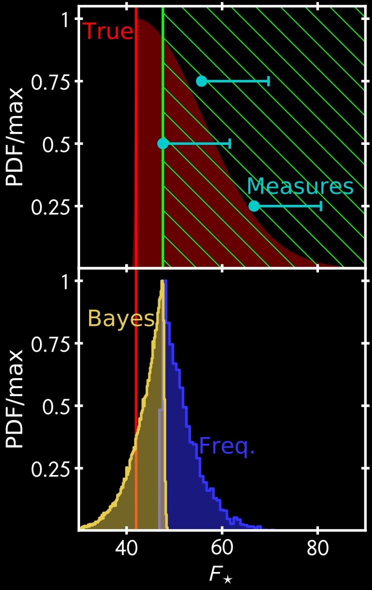
Noise: heavily-skewed split-normal distribution.

Common sense: true flux < minimum measure: $F_\star \lesssim 47.6$.

Frequentism: 95 % confidence interval: $[47.5, 62.5]$
$\Rightarrow$ inconsistent solution.

Bayesianism: 95 % credible range: $[34.6, 47.7]$
$\Rightarrow$ consistent solution.

This is an example of the *Jeffreys-Lindley's paradox*.



(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

Common sense: true flux < minimum measure: $F_\star \lesssim 47.6$.

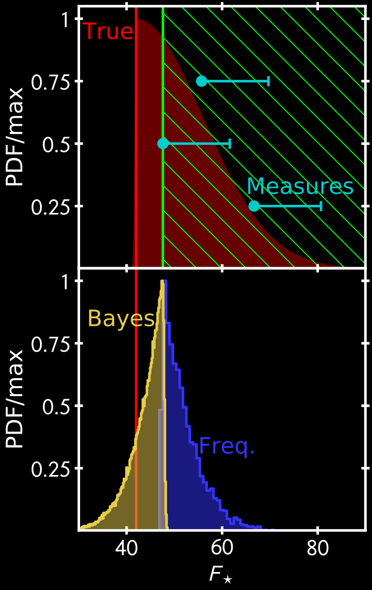
Frequentism: 95 % confidence interval: $[47.5, 62.5]$
$\Rightarrow$ inconsistent solution.

Bayesianism: 95 % credible range: $[34.6, 47.7]$
$\Rightarrow$ consistent solution.

This is an example of the *Jeffreys-Lindley's paradox*.

Difficulty of Interpreting Frequentist Results:

The frequentist interpretation leads to irrelevant interpretations



(Example adapted from Jaynes, 1976)

Difference Between the Two Approaches: Lindley's Paradox

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

Common sense: true flux < minimum measure: $F_\star \lesssim 47.6$.

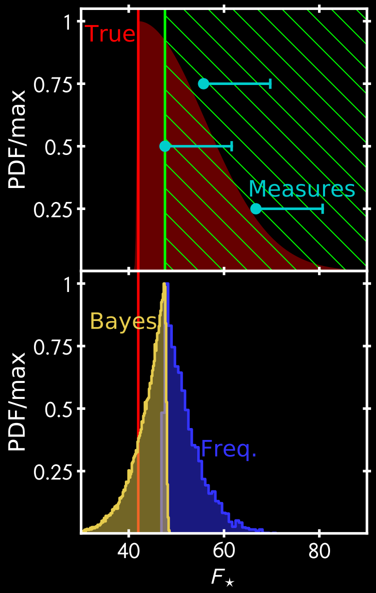
Frequentism: 95 % confidence interval: $[47.5, 62.5]$ $\Rightarrow$ inconsistent solution.

Bayesianism: 95 % credible range: $[34.6, 47.7]$ $\Rightarrow$ consistent solution.

This is an example of the *Jeffreys-Lindley's paradox*.

Difficulty of Interpreting Frequentist Results:

The frequentist interpretation leads to irrelevant interpretations $\Rightarrow$ question frequentist confidence intervals & p-values.



(Example adapted from Jaynes, 1976)

Same Problem with Asymmetric Noise:

Noise: heavily-skewed split-normal distribution.

Common sense: true flux < minimum measure: $F_\star \lesssim 47.6$.

Frequentism: 95 % confidence interval: $[47.5, 62.5]$
$\Rightarrow$ inconsistent solution.

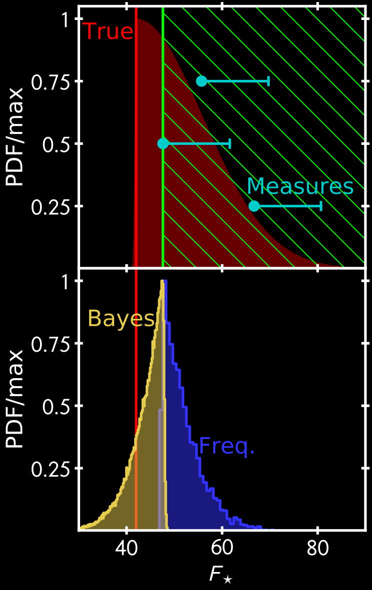
Bayesianism: 95 % credible range: $[34.6, 47.7]$
$\Rightarrow$ consistent solution.

This is an example of the *Jeffreys-Lindley's paradox*.

Difficulty of Interpreting Frequentist Results:

The frequentist interpretation leads to irrelevant interpretations $\Rightarrow$ question frequentist confidence intervals & p-values.

> *Bayesians address the question everyone is interested in by using assumptions no-one believes, while Frequentists use impeccable logic to deal with an issue of no interest to anyone.*
>
> (Lyons, 2013)



(Example adapted from Jaynes, 1976)

Hypothesis Testing: the Limitations of Frequentist p-Values

Bayesian Hypothesis Testing:

Hypothesis Testing: the Limitations of Frequentist p-Values

Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} =$$

Hypothesis Testing: the Limitations of Frequentist p-Values

Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Hypothesis Testing: the Limitations of Frequentist p-Values

Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

(Jeffreys, 1961)

Hypothesis Testing: the Limitations of Frequentist p-Values

Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
$>$ 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:

Hypothesis Testing: the Limitations of Frequentist p-Values

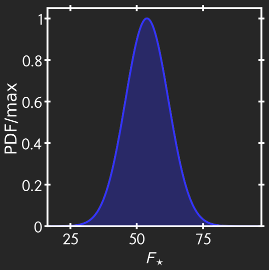
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
$>$ 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



Hypothesis Testing: the Limitations of Frequentist p-Values

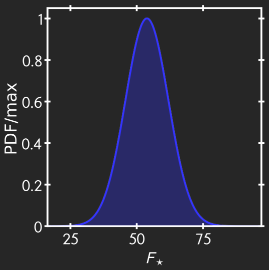
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



H_0: the *null hypothesis*;

Hypothesis Testing: the Limitations of Frequentist p-Values

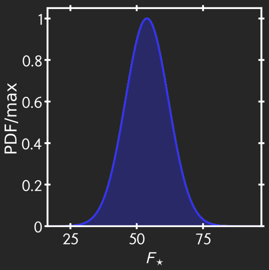
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
$>$ 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



H_0: the *null hypothesis*;

H_1: complement.

Hypothesis Testing: the Limitations of Frequentist p-Values

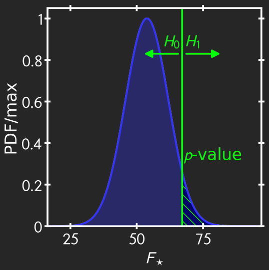
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
$>$ 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



H_0: the *null hypothesis*;

H_1: complement.

Hypothesis Testing: the Limitations of Frequentist p-Values

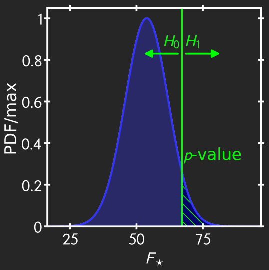
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



H_0: the *null hypothesis*;

H_1: complement.

Commonly:
$p < 0.05$
$\Rightarrow$ reject H_0.

Hypothesis Testing: the Limitations of Frequentist p-Values

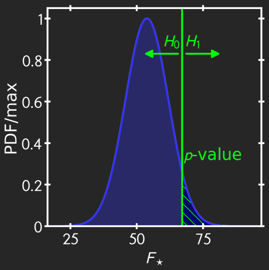
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



H_0: the *null hypothesis*;

H_1: complement.

Commonly:
$p < 0.05$
$\Rightarrow$ reject H_0.

Recent Controversy About the Interpretation & the Significance of p-Values:

Hypothesis Testing: the Limitations of Frequentist p-Values

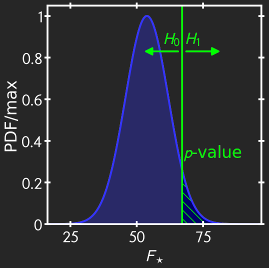
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



H_0: the *null hypothesis*;

H_1: complement.

Commonly:
$p < 0.05$
$\Rightarrow$ reject H_0.

Recent Controversy About the Interpretation & the Significance of p-Values:

2011: concept of *p-hacking* (Simmons *et al.*, 2011).

Hypothesis Testing: the Limitations of Frequentist p-Values

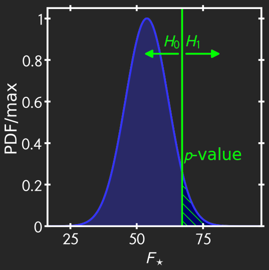
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
$>$ 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



H_0: the *null hypothesis*;

H_1: complement.

Commonly:
$p < 0.05$
$\Rightarrow$ reject H_0.

Recent Controversy About the Interpretation & the Significance of p-Values:

2011: concept of *p-hacking* (Simmons *et al.*, 2011).

2015: *Basic & Applied Social Psychology* *"would no longer publish papers containing p-values because they were too often used to support lower-quality research".*

Hypothesis Testing: the Limitations of Frequentist p-Values

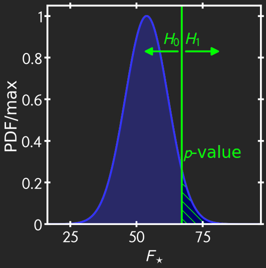
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
>100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



H_0: the *null hypothesis*;

H_1: complement.

Commonly:
$p < 0.05$
$\Rightarrow$ reject H_0.

Recent Controversy About the Interpretation & the Significance of p-Values:

2011: concept of *p-hacking* (Simmons *et al.*, 2011).

2015: *Basic & Applied Social Psychology* *“would no longer publish papers containing p-values because they were too often used to support lower-quality research”*.

2016: statement of the *American Statistical Association*: *“widespread use of 'statistical significance' (generally interpreted as 'p<0.05') as a license for making a claim of a scientific finding (or implied truth) leads to considerable distortion of the scientific process”*

Hypothesis Testing: the Limitations of Frequentist p-Values

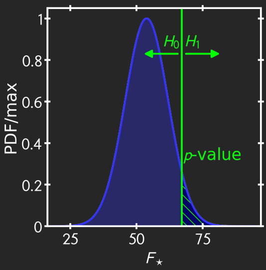
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
$>$ 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



H_0: the *null hypothesis*;

H_1: complement.

Commonly:
$p < 0.05$
$\Rightarrow$ reject H_0.

Recent Controversy About the Interpretation & the Significance of p-Values:

2011: concept of *p-hacking* (Simmons *et al.*, 2011).

2015: *Basic & Applied Social Psychology* *"would no longer publish papers containing p-values because they were too often used to support lower-quality research".*

2016: statement of the *American Statistical Association*: *"widespread use of 'statistical significance' (generally interpreted as 'p<0.05') as a license for making a claim of a scientific finding (or implied truth) leads to considerable distortion of the scientific process"* $\Rightarrow$ *"post p<0.05 era".*

Hypothesis Testing: the Limitations of Frequentist p-Values

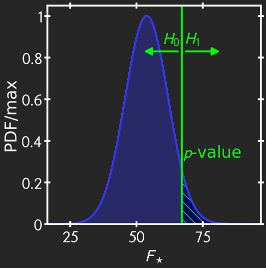
Bayesian Hypothesis Testing:

$$\underbrace{\frac{p(H_1|data)}{p(H_0|data)}}_{\text{posterior odds}} = \underbrace{\frac{p(data|H_1)}{p(data|H_0)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{prior odds}}$$

Bayes factor	Strength of evidence
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

(Jeffreys, 1961)

Frequentist Hypothesis Testing:



H_0: the *null hypothesis*;

H_1: complement.

Commonly:
$p < 0.05$
$\Rightarrow$ reject H_0.

Recent Controversy About the Interpretation & the Significance of p-Values:

2011: concept of *p-hacking* (Simmons *et al.*, 2011).

2015: *Basic & Applied Social Psychology* *"would no longer publish papers containing p-values because they were too often used to support lower-quality research".*

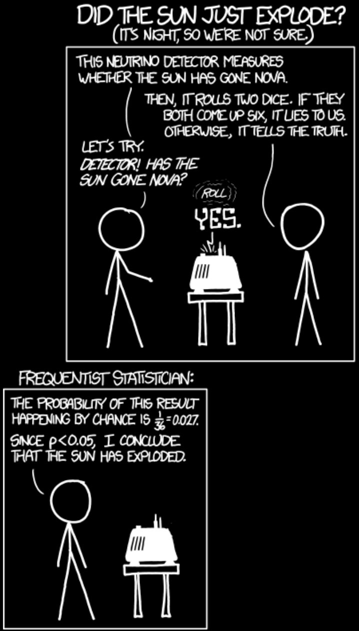
2016: statement of the *American Statistical Association*: *"widespread use of 'statistical significance' (generally interpreted as 'p<0.05') as a license for making a claim of a scientific finding (or implied truth) leads to considerable distortion of the scientific process"* $\Rightarrow$ *"post p<0.05 era".*

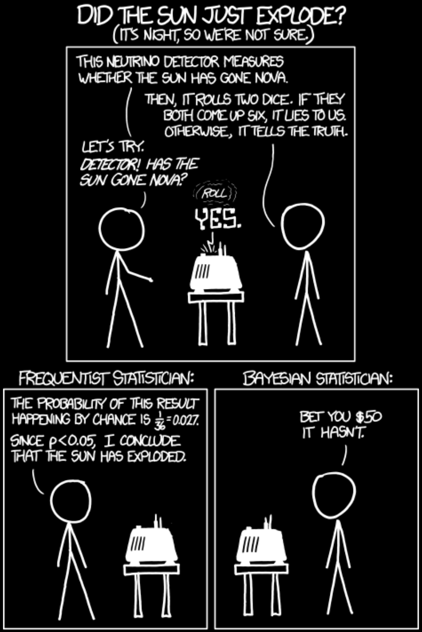
2018: *Political Analysis* *"will no longer be reporting p-values".*

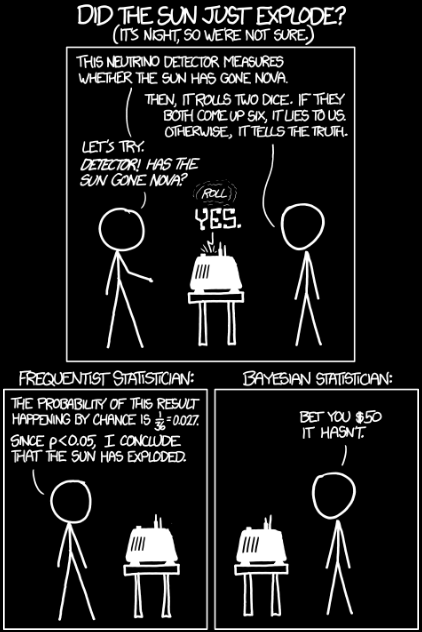
Reductio Ad Absurdum (xkcd)


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
(ROLL)
YES.

Reductio Ad Absurdum (xkcd)



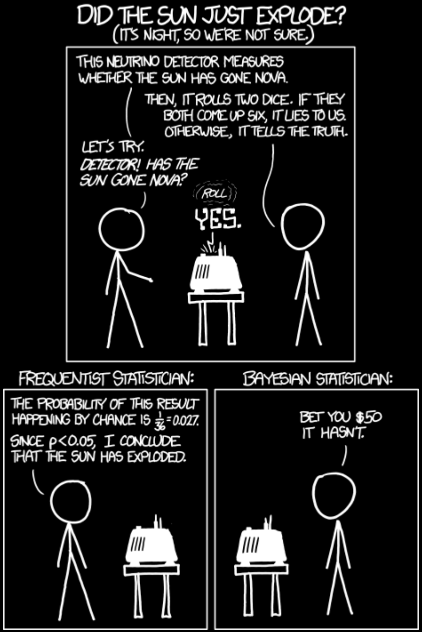

DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS 1/36 = 0.027. SINCE p<0.05, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.



The Bayesian Point of View:

$$p(\text{nova}|2 \times 6) = \frac{p(2 \times 6|\text{nova}) \times p(\text{nova})}{p(2 \times 6)}$$

Reductio Ad Absurdum (xkcd)



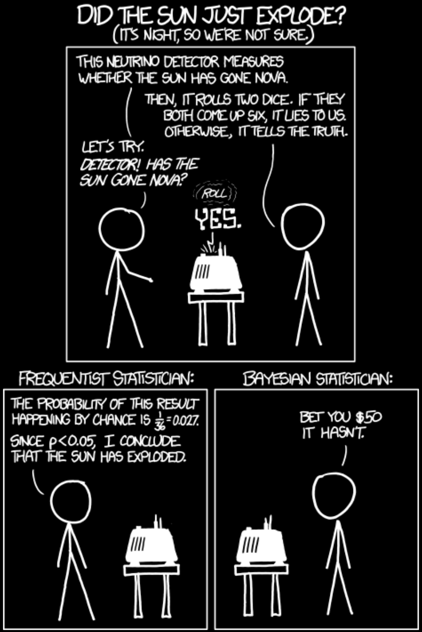
The Bayesian Point of View:

$$p(\text{nova}|2\times 6) = \frac{p(2\times 6|\text{nova})\times p(\text{nova})}{p(2\times 6)}$$

Laplace's Law of Succession:

(Essai philosophique sur les probabilités, 1814)

Reductio Ad Absurdum (xkcd)



The Bayesian Point of View:

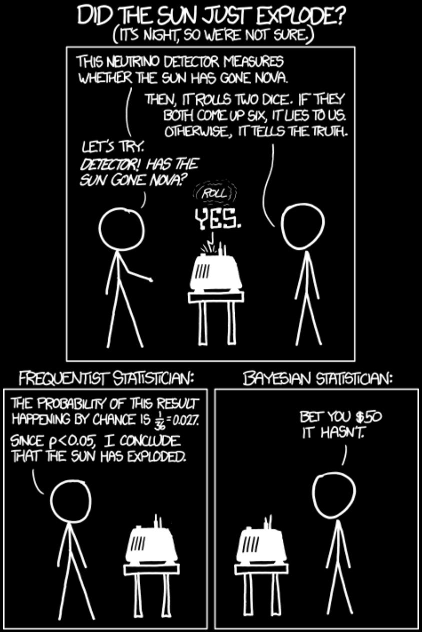
$$p(\text{nova}|2\times 6) = \frac{p(2\times 6|\text{nova}) \times p(\text{nova})}{p(2\times 6)}$$

Laplace's Law of Succession:

- Probability of an event, knowing it happened n consecutive times $= (n+1)/(n+2)$.

(Essai philosophique sur les probabilités, 1814)

Reductio Ad Absurdum (xkcd)



The Bayesian Point of View:

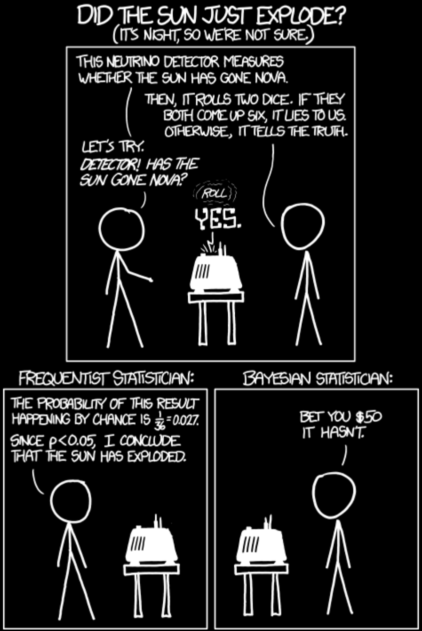
$$p(\text{nova}|2\times 6) = \frac{p(2\times 6|\text{nova})\times p(\text{nova})}{p(2\times 6)}$$

Laplace's Law of Succession:

- Probability of an event, knowing it happened n consecutive times $= (n+1)/(n+2)$.
- The Sun is 5 000 years $\Rightarrow n = 1\,826\,213$ (Bible).

(Essai philosophique sur les probabilités, 1814)

Reductio Ad Absurdum (xkcd)



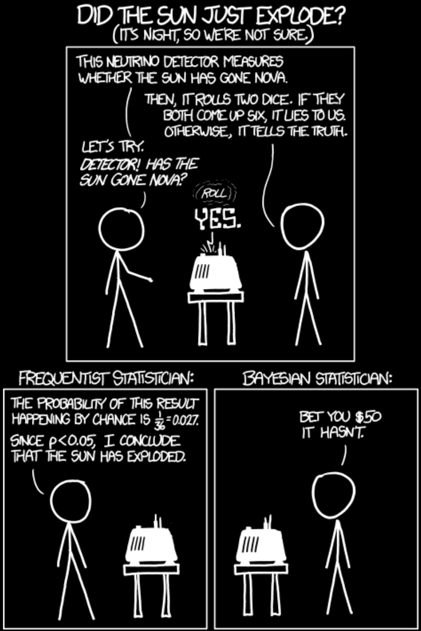
The Bayesian Point of View:

$$p(\text{nova}|2\times 6) = \frac{p(2\times 6|\text{nova})\times p(\text{nova})}{p(2\times 6)}$$

Laplace's Law of Succession:

- Probability of an event, knowing it happened n consecutive times $= (n+1)/(n+2)$.
- The Sun is 5 000 years $\Rightarrow n = 1\,826\,213$ (Bible).
- $p(\text{nova}) = 1 - p(\text{sunrise}) = 1/1\,826\,215$.

(Essai philosophique sur les probabilités, 1814)



The Bayesian Point of View:

$$p(\text{nova}|2\times 6) = \frac{p(2\times 6|\text{nova}) \times p(\text{nova})}{p(2\times 6)}$$

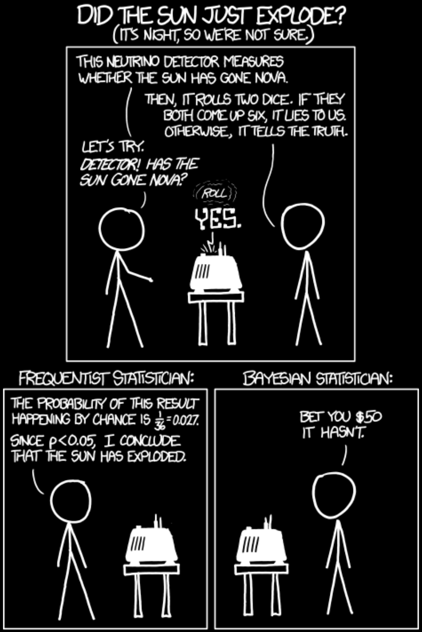
Laplace's Law of Succession:

- Probability of an event, knowing it happened n consecutive times $= (n+1)/(n+2)$.
- The Sun is 5 000 years $\Rightarrow n = 1\,826\,213$ (Bible).
- $p(\text{nova}) = 1 - p(\text{sunrise}) = 1/1\,826\,215$.

(Essai philosophique sur les probabilités, 1814)

The Bayesian Solution:

Reductio Ad Absurdum (xkcd)



The Bayesian Point of View:

$$p(\text{nova}|2\times 6) = \frac{p(2\times 6|\text{nova})\times p(\text{nova})}{p(2\times 6)}$$

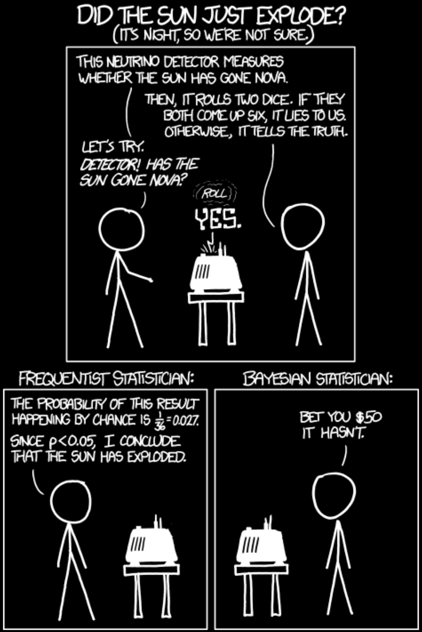
Laplace's Law of Succession:

- Probability of an event, knowing it happened n consecutive times $= (n+1)/(n+2)$.
- The Sun is 5 000 years $\Rightarrow n = 1\,826\,213$ (Bible).
- $p(\text{nova}) = 1 - p(\text{sunrise}) = 1/1\,826\,215$.

(Essai philosophique sur les probabilités, 1814)

The Bayesian Solution:

$$p(2\times 6) \quad = \quad 1/36$$



The Bayesian Point of View:

$$p(\text{nova}|2\times 6) = \frac{p(2\times 6|\text{nova}) \times p(\text{nova})}{p(2\times 6)}$$

Laplace's Law of Succession:

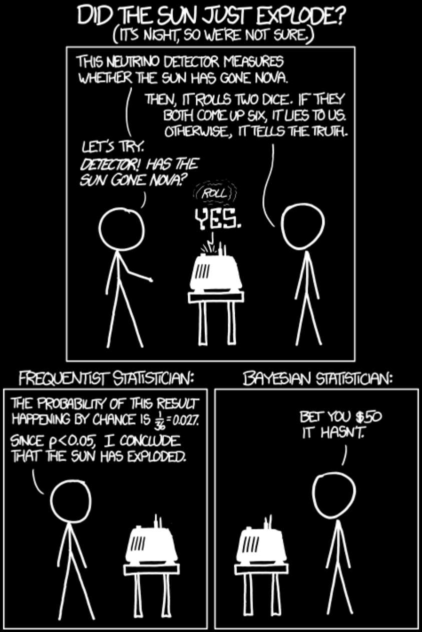
- Probability of an event, knowing it happened n consecutive times $= (n+1)/(n+2)$.
- The Sun is 5 000 years $\Rightarrow n = 1\,826\,213$ (Bible).
- $p(\text{nova}) = 1 - p(\text{sunrise}) = 1/1\,826\,215$.

(Essai philosophique sur les probabilités, 1814)

The Bayesian Solution:

$$\begin{aligned} p(2\times 6) &= 1/36 \\ p(2\times 6|\text{nova}) &= 1 \end{aligned}$$

Reductio Ad Absurdum (xkcd)



The Bayesian Point of View:

$$p(\text{nova}|2\times 6) = \frac{p(2\times 6|\text{nova})\times p(\text{nova})}{p(2\times 6)}$$

Laplace's Law of Succession:

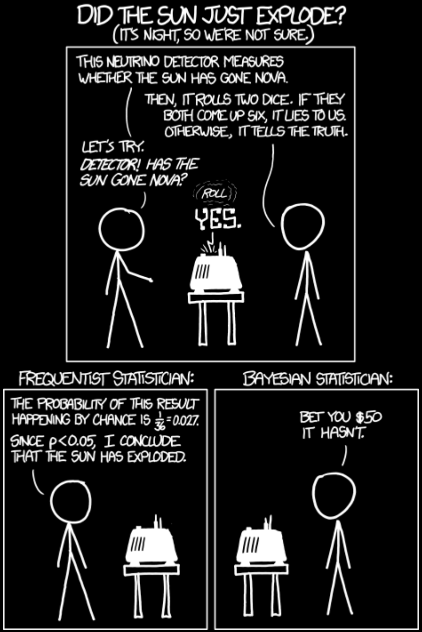
- Probability of an event, knowing it happened n consecutive times $= (n+1)/(n+2)$.
- The Sun is 5 000 years $\Rightarrow n = 1\,826\,213$ (Bible).
- $p(\text{nova}) = 1 - p(\text{sunrise}) = 1/1\,826\,215$.

(Essai philosophique sur les probabilités, 1814)

The Bayesian Solution:

$$\begin{aligned} p(2\times 6) &= 1/36 \\ p(2\times 6|\text{nova}) &= 1 \\ p(\text{nova}) &= 1/1\,826\,215 \end{aligned}$$

Reductio Ad Absurdum (xkcd)



The Bayesian Point of View:

$$p(\text{nova}|2\times 6) = \frac{p(2\times 6|\text{nova}) \times p(\text{nova})}{p(2\times 6)}$$

Laplace's Law of Succession:

- Probability of an event, knowing it happened n consecutive times $= (n+1)/(n+2)$.
- The Sun is 5 000 years $\Rightarrow$ $n = 1\,826\,213$ (Bible).
- $p(\text{nova}) = 1 - p(\text{sunrise}) = 1/1\,826\,215$.

(Essai philosophique sur les probabilités, 1814)

The Bayesian Solution:

$$\begin{aligned} p(2\times 6) &= 1/36 \\ p(2\times 6|\text{nova}) &= 1 \\ p(\text{nova}) &= 1/1\,826\,215 \\ \Rightarrow p(\text{nova}|2\times 6) &\simeq 2\times 10^{-5} \end{aligned}$$

The Bayesian Approach is Holistic: Stein's Paradox

The Bayesian Approach is Holistic: Stein's Paradox

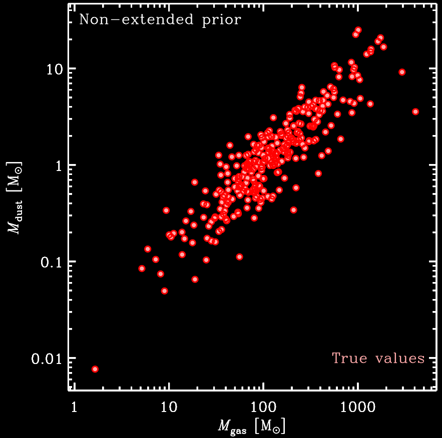
Prior Depends on Dust Model Parameters:

Intuitive approach: $p(M_{\text{dust}})$.

The Bayesian Approach is Holistic: Stein's Paradox

Prior Depends on Dust Model Parameters:

Intuitive approach: $p(M_{\rm dust})$.

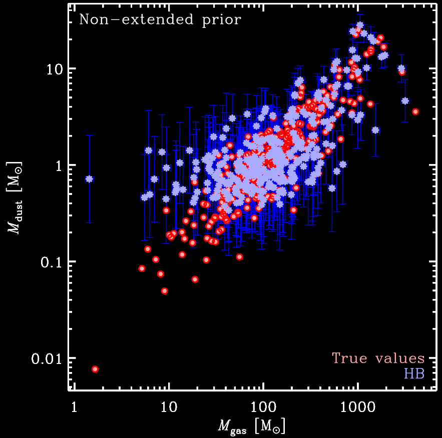


(Galliano, 2018)

The Bayesian Approach is Holistic: Stein's Paradox

Prior Depends on Dust Model Parameters:

Intuitive approach: $p(M_{\rm dust})$.



(Galliano, 2018)

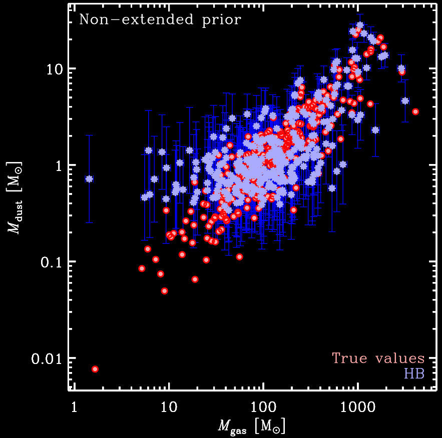
The Bayesian Approach is Holistic: Stein's Paradox

Prior Depends on Dust Model Parameters:

Intuitive approach: $p(M_{\rm dust})$.

Prior Also Includes Ancillary Data:

Holistic approach: $p(M_{\rm dust}, M_{\rm gas})$.



(Galliano, 2018)

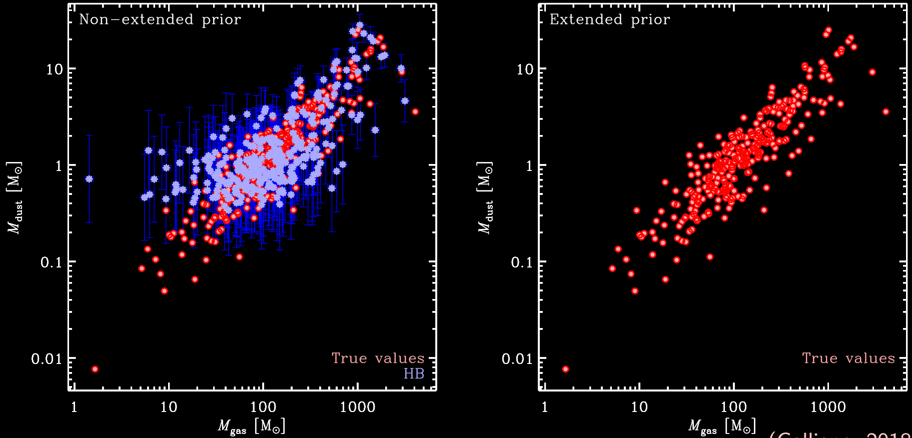
The Bayesian Approach is Holistic: Stein's Paradox

Prior Depends on Dust Model Parameters:

Intuitive approach: $p(M_{\rm dust})$.

Prior Also Includes Ancillary Data:

Holistic approach: $p(M_{\rm dust}, M_{\rm gas})$.



(Galliano, 2018)

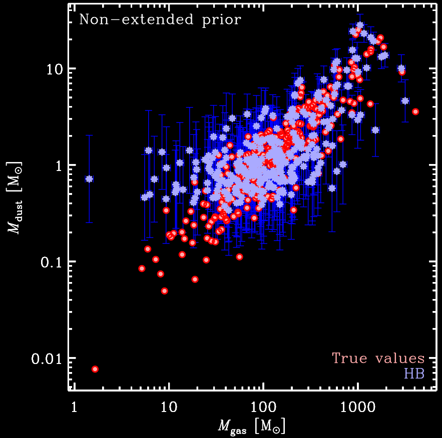
The Bayesian Approach is Holistic: Stein's Paradox

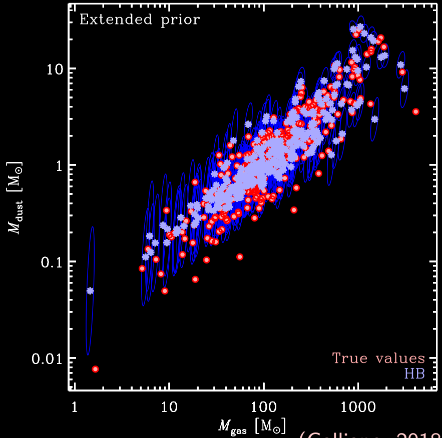
Prior Depends on Dust Model Parameters:

Intuitive approach: $p(M_{\rm dust})$.

Prior Also Includes Ancillary Data:

Holistic approach: $p(M_{\rm dust}, M_{\rm gas})$.





(Galliano, 2018)

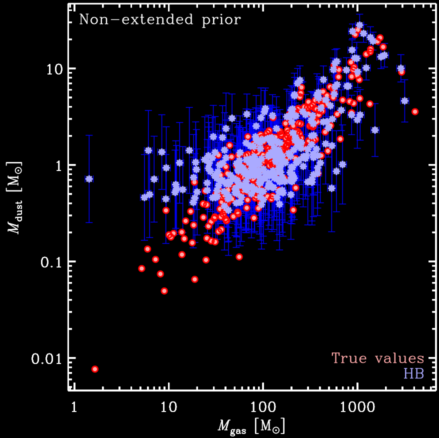
The Bayesian Approach is Holistic: Stein's Paradox

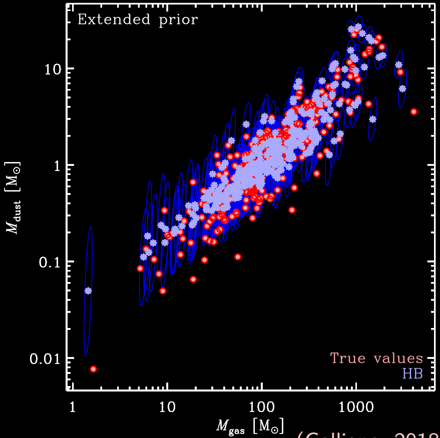
Prior Depends on Dust Model Parameters:

Intuitive approach: $p(M_{\text{dust}})$.

Prior Also Includes Ancillary Data:

Holistic approach: $p(M_{\text{dust}}, M_{\text{gas}})$.





(Galliano, 2018)

$\Rightarrow$ partition of knowledge is statistically inadmissible (Bayesian take on *Stein's paradox*).

Pros & Cons of the Two Approaches

Pros & Cons of the Two Approaches

The Bayesian Approach

The Frequentist Approach

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

CON choice of prior is arbitrary.

The Frequentist Approach

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

CON choice of prior is arbitrary.

The Frequentist Approach

PRO likelihood is not arbitrary.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

CON choice of prior is arbitrary.

PRO the posterior makes sense (conditional on the data) & is easy to interpret.

The Frequentist Approach

PRO likelihood is not arbitrary.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

CON choice of prior is arbitrary.

PRO the posterior makes sense (conditional on the data) & is easy to interpret.

The Frequentist Approach

PRO likelihood is not arbitrary.

CON samples non-observed data & arbitrary choice of estimator, loss functions or p-value.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.
- **CON** heavy computation.

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.
- **CON** heavy computation.

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.
- **PRO** fast computation.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.
- **CON** heavy computation.
- **PRO** works well with small samples, even with 0 data...

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.
- **PRO** fast computation.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.
- **CON** heavy computation.
- **PRO** works well with small samples, even with 0 data...

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.
- **PRO** fast computation.
- **CON** does not work well with small samples.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.
- **CON** heavy computation.
- **PRO** works well with small samples, even with 0 data...
- **PRO** master equation $\Rightarrow$ easier to teach.

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.
- **PRO** fast computation.
- **CON** does not work well with small samples.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.
- **CON** heavy computation.
- **PRO** works well with small samples, even with 0 data...
- **PRO** master equation $\Rightarrow$ easier to teach.

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.
- **PRO** fast computation.
- **CON** does not work well with small samples.
- **CON** difficult to teach (collection of *ad hoc* cooking recipes).

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.
- **CON** heavy computation.
- **PRO** works well with small samples, even with 0 data...
- **PRO** master equation $\Rightarrow$ easier to teach.
- **PRO** holistic & flexible: can account for all data & theories.

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.
- **PRO** fast computation.
- **CON** does not work well with small samples.
- **CON** difficult to teach (collection of *ad hoc* cooking recipes).

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.
- **CON** heavy computation.
- **PRO** works well with small samples, even with 0 data...
- **PRO** master equation $\Rightarrow$ easier to teach.
- **PRO** holistic & flexible: can account for all data & theories.

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.
- **PRO** fast computation.
- **CON** does not work well with small samples.
- **CON** difficult to teach (collection of *ad hoc* cooking recipes).
- **CON** strict: can account only for data related to a particular experiment.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.
- **CON** heavy computation.
- **PRO** works well with small samples, even with 0 data...
- **PRO** master equation $\Rightarrow$ easier to teach.
- **PRO** holistic & flexible: can account for all data & theories.
- **PRO** conservative.

The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.
- **PRO** fast computation.
- **CON** does not work well with small samples.
- **CON** difficult to teach (collection of *ad hoc* cooking recipes).
- **CON** strict: can account only for data related to a particular experiment.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Pros & Cons of the Two Approaches

The Bayesian Approach

- **CON** choice of prior is arbitrary.
- **PRO** the posterior makes sense (conditional on the data) & is easy to interpret.
- **PRO** probabilistic logic $\Rightarrow$ continuum between skepticism & confidence.
- **CON** heavy computation.
- **PRO** works well with small samples, even with 0 data...
- **PRO** master equation $\Rightarrow$ easier to teach.
- **PRO** holistic & flexible: can account for all data & theories.
- **PRO** conservative.

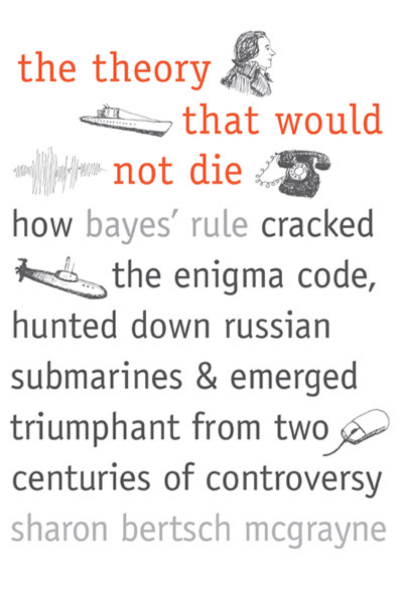
The Frequentist Approach

- **PRO** likelihood is not arbitrary.
- **CON** samples non-observed data & arbitrary choice of estimator, loss functions or p-value.
- **CON** boolean logic $\Rightarrow$ a proposition is either true or false.
- **PRO** fast computation.
- **CON** does not work well with small samples.
- **CON** difficult to teach (collection of *ad hoc* cooking recipes).
- **CON** strict: can account only for data related to a particular experiment.
- **CON** can give ridiculous answers.

(*cf.* reviews by *e.g.* Jaynes, 1976; Loredo, 1990; Lindley, 1999; Wagenmacker, 2008; Lyons, 2013)

Outline of the Talk

A Comprehensive Historical Perspective: S. McGrayne's Book



Sharon Bertsch McGRAYNE (1942–)

(Published in 2011)

First Formulation by Bayes

First Formulation by Bayes

Thomas BAYES
($\simeq$1701–1761)

First Formulation by Bayes

Thomas BAYES
($\simeq$1701–1761)

Intellectual Interests:

First Formulation by Bayes

Thomas BAYES
(≃1701–1761)

Intellectual Interests:

- Presbyterian reverend from a *nonconformist* family.

First Formulation by Bayes

Thomas BAYES
(≃1701–1761) 🇬🇧

Intellectual Interests:

- Presbyterian reverend from a *nonconformist* family.
- Read: (1) David HUME (philosophy); (2) Isaac NEWTON (physics); and (3) Abraham DE MOIVRE (proabilities).

First Formulation by Bayes

Thomas BAYES
(≃1701–1761)

Intellectual Interests:

- Presbyterian reverend from a *nonconformist* family.
- Read: (1) David HUME (philosophy); (2) Isaac NEWTON (physics); and (3) Abraham DE MOIVRE (proabilities).
- Published, during his lifetime: (1) a treaty of theology; and (2) a treaty of mathematics.

First Formulation by Bayes

Thomas BAYES
(≃1701–1761)

Intellectual Interests:

- Presbyterian reverend from a *nonconformist* family.
- Read: (1) David HUME (philosophy); (2) Isaac NEWTON (physics); and (3) Abraham DE MOIVRE (proabilities).
- Published, during his lifetime: (1) a treaty of theology; and (2) a treaty of mathematics.

The Discovery of Bayes' Rule:

First Formulation by Bayes

Thomas BAYES
(≃1701–1761)

Intellectual Interests:

- Presbyterian reverend from a *nonconformist* family.
- Read: (1) David HUME (philosophy); (2) Isaac NEWTON (physics); and (3) Abraham DE MOIVRE (proabilities).
- Published, during his lifetime: (1) a treaty of theology; and (2) a treaty of mathematics.

The Discovery of Bayes' Rule:

Thought Experiment: between 1746 and 1749, he formulates his rule to infer the position of a ball on a pool table. Prior: position relative to another ball.

First Formulation by Bayes

Thomas BAYES
($\simeq$1701–1761)

Intellectual Interests:

- Presbyterian reverend from a *nonconformist* family.
- Read: (1) David HUME (philosophy); (2) Isaac NEWTON (physics); and (3) Abraham DE MOIVRE (proabilities).
- Published, during his lifetime: (1) a treaty of theology; and (2) a treaty of mathematics.

The Discovery of Bayes' Rule:

Thought Experiment: between 1746 and 1749, he formulates his rule to infer the position of a ball on a pool table. Prior: position relative to another ball.

Probability of an event: *"the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening"* (An Essay towards solving a Problem in the Doctrine of Chances, 1764).

First Formulation by Bayes

Thomas BAYES
(≃1701–1761)

Intellectual Interests:

- Presbyterian reverend from a *nonconformist* family.
- Read: (1) David HUME (philosophy); (2) Isaac NEWTON (physics); and (3) Abraham DE MOIVRE (proabilities).
- Published, during his lifetime: (1) a treaty of theology; and (2) a treaty of mathematics.

The Discovery of Bayes' Rule:

Thought Experiment: between 1746 and 1749, he formulates his rule to infer the position of a ball on a pool table. Prior: position relative to another ball.

Probability of an event: *"the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening"* (An Essay towards solving a Problem in the Doctrine of Chances, 1764).

Publication: in 1764, by Richard PRICE: his formula gives the probability of causes ⇒ can be applied to prove God's existence.

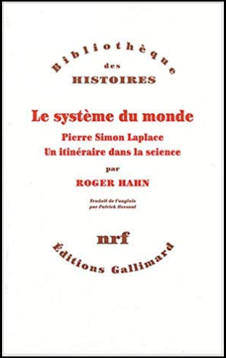
Laplace: the Probability of Causes of Events (1/2)

Laplace: the Probability of Causes of Events (1/2)

Pierre Simon LAPLACE (1749–1827)

Laplace: the Probability of Causes of Events (1/2)

Pierre Simon LAPLACE (1749–1827)



Bibliothèque des HISTOIRES

Le système du monde

Pierre Simon Laplace

Un itinéraire dans la science

par

ROGER HAHN

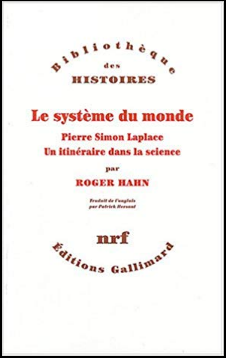
nrf

Éditions Gallimard

(Hahn, 2004; English version in 2005)

Laplace: the Probability of Causes of Events (1/2)

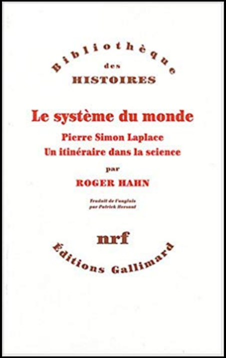
Pierre Simon LAPLACE (1749–1827)



(Hahn, 2004; English version in 2005)

Upbringing & Early Career:

Pierre Simon LAPLACE (1749–1827) 🇫🇷

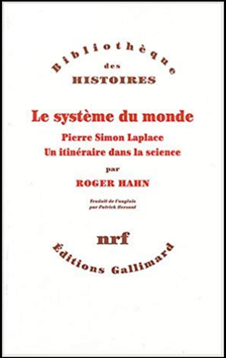


(Hahn, 2004; English version in 2005)

Upbringing & Early Career:

- He Grew-up near Caen (Normandie), son of a cider producer & small estate owner.

Pierre Simon LAPLACE (1749–1827)



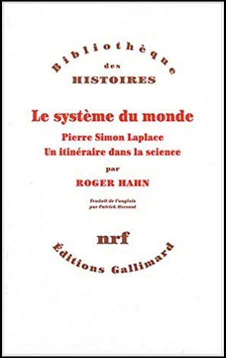
(Hahn, 2004; English version in 2005)

Upbringing & Early Career:

- He Grew-up near Caen (Normandie), son of a cider producer & small estate owner.
- His father pushes him towards a religious career $\Rightarrow$ studies theology at the University of Caen.

Laplace: the Probability of Causes of Events (1/2)

Pierre Simon LAPLACE (1749–1827)

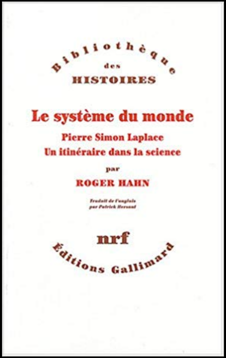


(Hahn, 2004; English version in 2005)

Upbringing & Early Career:

- He Grew-up near Caen (Normandie), son of a cider producer & small estate owner.
- His father pushes him towards a religious career ⇒ studies theology at the University of Caen.
- Abandons this idea and moves to Paris ⇒ meets d'Alembert in 1769, who helps him to get a teaching position at the *École Royale Militaire*.

Pierre Simon LAPLACE (1749–1827)



(Hahn, 2004; English version in 2005)

Upbringing & Early Career:

- He Grew-up near Caen (Normandie), son of a cider producer & small estate owner.
- His father pushes him towards a religious career ⇒ studies theology at the University of Caen.
- Abandons this idea and moves to Paris ⇒ meets d'Alembert in 1769, who helps him to get a teaching position at the *École Royale Militaire*.
- Applies to the *Académie Royale des Sciences* ⇒ elected in 1773.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir $\Rightarrow$ understood probabilities can be used to quantify observational errors.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir $\Rightarrow$ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir $\Rightarrow$ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) $\Rightarrow$ should say *Bayesian-Laplacian method*.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir $\Rightarrow$ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) $\Rightarrow$ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir $\Rightarrow$ discovered it in 1781, when Price went to Paris.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir $\Rightarrow$ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) $\Rightarrow$ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir $\Rightarrow$ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir $\Rightarrow$ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) $\Rightarrow$ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir $\Rightarrow$ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography $\Rightarrow$ birth rate of males higher than females is a general rule of human kind.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir $\Rightarrow$ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) $\Rightarrow$ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir $\Rightarrow$ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography $\Rightarrow$ birth rate of males higher than females is a general rule of human kind.

(...) la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

(Théorie analytique des probabilités, 1812)

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir ⇒ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) ⇒ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir ⇒ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography ⇒ birth rate of males higher than females is a general rule of human kind.

(...) la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

(Théorie analytique des probabilités, 1812)

Achievements in Celestial Mechanics:

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir $\Rightarrow$ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) $\Rightarrow$ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir $\Rightarrow$ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography $\Rightarrow$ birth rate of males higher than females is a general rule of human kind.

(...) la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

(Théorie analytique des probabilités, 1812)

Achievements in Celestial Mechanics:

- Newton's geometric approach of mechanics $\Rightarrow$ analytic approach.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir ⇒ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) ⇒ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir ⇒ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography ⇒ birth rate of males higher than females is a general rule of human kind.

(...) la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

(Théorie analytique des probabilités, 1812)

Achievements in Celestial Mechanics:

- Newton's geometric approach of mechanics ⇒ analytic approach.
- Origin & stability of the Solar system.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir ⇒ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) ⇒ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir ⇒ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography ⇒ birth rate of males higher than females is a general rule of human kind.

(...) la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

(Théorie analytique des probabilités, 1812)

Achievements in Celestial Mechanics:

- Newton's geometric approach of mechanics ⇒ analytic approach.
- Origin & stability of the Solar system.
- Postulated existence of black holes.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir ⇒ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) ⇒ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir ⇒ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography ⇒ birth rate of males higher than females is a general rule of human kind.

(...) la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

(Théorie analytique des probabilités, 1812)

Achievements in Celestial Mechanics:

- Newton's geometric approach of mechanics ⇒ analytic approach.
- Origin & stability of the Solar system.
- Postulated existence of black holes.

Political Career:

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir ⇒ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) ⇒ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir ⇒ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography ⇒ birth rate of males higher than females is a general rule of human kind.

(...) la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

(Théorie analytique des probabilités, 1812)

Achievements in Celestial Mechanics:

- Newton's geometric approach of mechanics ⇒ analytic approach.
- Origin & stability of the Solar system.
- Postulated existence of black holes.

Political Career:

1799: (republic) interior minister.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir ⇒ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) ⇒ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir ⇒ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography ⇒ birth rate of males higher than females is a general rule of human kind.

(...) la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

(Théorie analytique des probabilités, 1812)

Achievements in Celestial Mechanics:

- Newton's geometric approach of mechanics ⇒ analytic approach.
- Origin & stability of the Solar system.
- Postulated existence of black holes.

Political Career:

1799: (republic) interior minister.

1808: (empire) Comte d'Empire.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir $\Rightarrow$ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) $\Rightarrow$ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir $\Rightarrow$ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography $\Rightarrow$ birth rate of males higher than females is a general rule of human kind.

(...) la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

(Théorie analytique des probabilités, 1812)

Achievements in Celestial Mechanics:

- Newton's geometric approach of mechanics $\Rightarrow$ analytic approach.
- Origin & stability of the Solar system.
- Postulated existence of black holes.

Political Career:

1799: (republic) interior minister.

1808: (empire) Comte d'Empire.

1817: (monarchy) Marquis.

Laplace: the Probability of Causes of Events (2/2)

Contribution to the Theory of Probabilities:

- Read Abraham DE MOIVRE's memoir $\Rightarrow$ understood probabilities can be used to quantify observational errors.
- Wrote: Mémoire sur la probabilité des causes par les événements (1774), where he rediscovered Bayes' rule.
  - First application of Bayes' rule (law of succession) $\Rightarrow$ should say *Bayesian-Laplacian method*.
  - He had not read Bayes memoir $\Rightarrow$ discovered it in 1781, when Price went to Paris.
  - Laplace acknowledged Bayes.
- Studied demography $\Rightarrow$ birth rate of males higher than females is a general rule of human kind.

(...) la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte.

(Théorie analytique des probabilités, 1812)

Achievements in Celestial Mechanics:

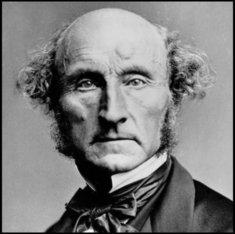
- Newton's geometric approach of mechanics $\Rightarrow$ analytic approach.
- Origin & stability of the Solar system.
- Postulated existence of black holes.

Political Career:

1799: (republic) interior minister.
1808: (empire) Comte d'Empire.
1817: (monarchy) Marquis.
1814-1827: member of parliament.

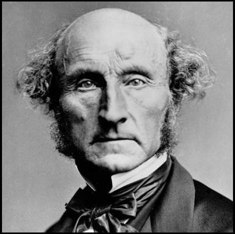
After Laplace: Early Frequentist Development

After Laplace: Early Frequentist Development

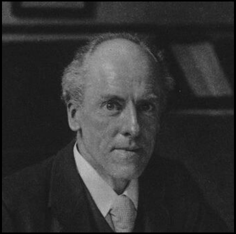


John Stuart MILL
(1806–1873)

After Laplace: Early Frequentist Development

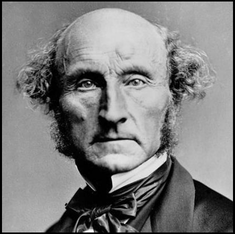


John Stuart MILL
(1806–1873)

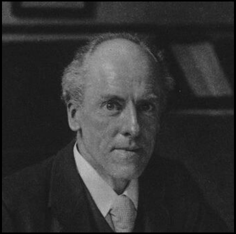


Karl PEARSON
(1857–1936)

After Laplace: Early Frequentist Development



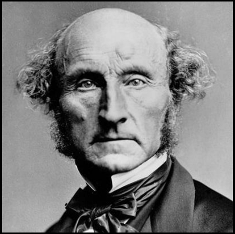
John Stuart MILL
(1806–1873)



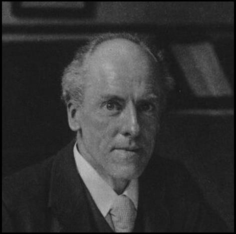
Karl PEARSON
(1857–1936)

Rejection of Laplace's Work:

After Laplace: Early Frequentist Development



John Stuart MILL
(1806–1873)

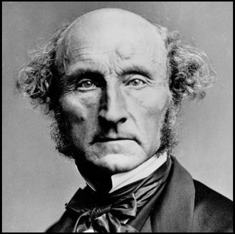


Karl PEARSON
(1857–1936)

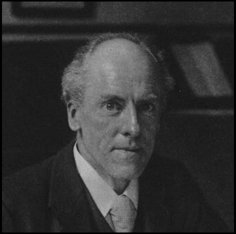
Rejection of Laplace's Work:

Mill initiated the frequentist approach 10 years after Laplace's death.

After Laplace: Early Frequentist Development



John Stuart MILL
(1806–1873)



Karl PEARSON
(1857–1936)

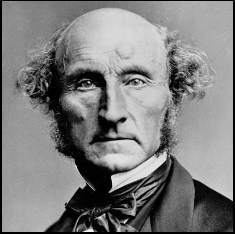
Rejection of Laplace's Work:

Mill initiated the frequentist approach 10 years after Laplace's death.

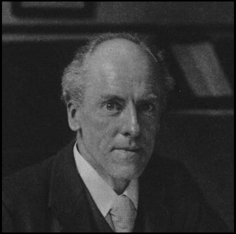
> *(...) a very slight improvement in the data, by better observations, or by taking into fuller consideration the special circumstances of the case, is of more use than the most elaborate application of the calculus of probabilities founded on the data in their previous state of inferiority. The neglect of this obvious reflection has given rise to misapplications of the calculus of probabilities which have made it the real opprobrium of mathematics.*

(A System of Logic, 1843)

After Laplace: Early Frequentist Development



John Stuart MILL
(1806–1873)



Karl PEARSON
(1857–1936)

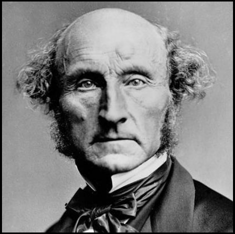
Rejection of Laplace's Work:

Mill initiated the frequentist approach 10 years after Laplace's death.

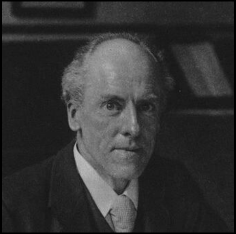
> *(...) a very slight improvement in the data, by better observations, or by taking into fuller consideration the special circumstances of the case, is of more use than the most elaborate application of the calculus of probabilities founded on the data in their previous state of inferiority. The neglect of this obvious reflection has given rise to misapplications of the calculus of probabilities which have made it the real opprobrium of mathematics.*
>
> (A System of Logic, 1843)

1. The idea that probability should represent a degree of plausability seemed too vague to be the foundation for a mathematical theory & no clear way to assign priors.

After Laplace: Early Frequentist Development



John Stuart MILL
(1806–1873)



Karl PEARSON
(1857–1936)

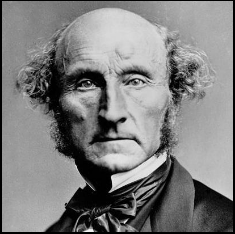
Rejection of Laplace's Work:

Mill initiated the frequentist approach 10 years after Laplace's death.

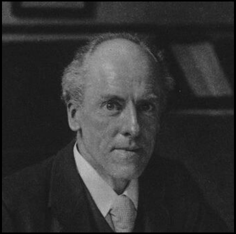
> *(...) a very slight improvement in the data, by better observations, or by taking into fuller consideration the special circumstances of the case, is of more use than the most elaborate application of the calculus of probabilities founded on the data in their previous state of inferiority. The neglect of this obvious reflection has given rise to misapplications of the calculus of probabilities which have made it the real opprobrium of mathematics.*
>
> (A System of Logic, 1843)

1. The idea that probability should represent a degree of plausability seemed too vague to be the foundation for a mathematical theory & no clear way to assign priors.
2. The computation of Bayesian solutions was difficult (no computers).

After Laplace: Early Frequentist Development



John Stuart MILL
(1806–1873)



Karl PEARSON
(1857–1936)

Rejection of Laplace's Work:

Mill initiated the frequentist approach 10 years after Laplace's death.

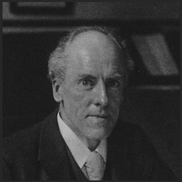
> *(...) a very slight improvement in the data, by better observations, or by taking into fuller consideration the special circumstances of the case, is of more use than the most elaborate application of the calculus of probabilities founded on the data in their previous state of inferiority. The neglect of this obvious reflection has given rise to misapplications of the calculus of probabilities which have made it the real opprobrium of mathematics.*
>
> (A System of Logic, 1843)

1. The idea that probability should represent a degree of plausability seemed too vague to be the foundation for a mathematical theory & no clear way to assign priors.
2. The computation of Bayesian solutions was difficult (no computers).
3. Laplace was despised in England for his *Bonapartism*.

The Golden Age of Frequentism (1920–1930)

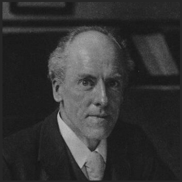
The Golden Age of Frequentism (1920–1930)

Karl PEARSON (1857–1936)



The Golden Age of Frequentism (1920–1930)

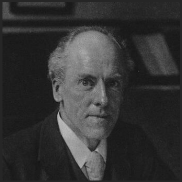
Karl PEARSON (1857–1936)



- He was a social darwinist & eugenist.

The Golden Age of Frequentism (1920–1930)

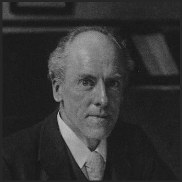
Karl PEARSON (1857–1936)



- He was a social darwinist & eugenist.
- The Grammar of Science (1892) influenced Einstein.

The Golden Age of Frequentism (1920–1930)

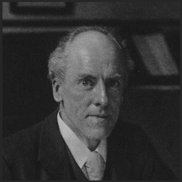
Karl PEARSON (1857–1936)



- He was a social darwinist & eugenist.
- The Grammar of Science (1892) influenced Einstein.
- Developed: (1) χ^2 test; (2) standard-deviation; (3) correlation; (4) *p*-value; (5) P.C.A.

The Golden Age of Frequentism (1920–1930)

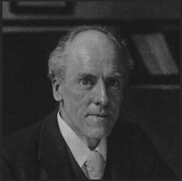
Karl PEARSON (1857–1936)



- He was a social darwinist & eugenist.
- The Grammar of Science (1892) influenced Einstein.
- Developed: (1) χ^2 test; (2) standard-deviation; (3) correlation; (4) *p*-value; (5) P.C.A.
- Initiator of the anti-Bayesian current.

The Golden Age of Frequentism (1920–1930)

Karl PEARSON (1857–1936)

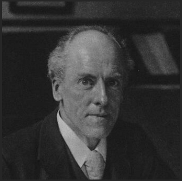


- He was a social darwinist & eugenist.
- The Grammar of Science (1892) influenced Einstein.
- Developed: (1) χ^2 test; (2) standard-deviation; (3) correlation; (4) p-value; (5) P.C.A.
- Initiator of the anti-Bayesian current.

Sir Ronald FISHER (1890–1962)

The Golden Age of Frequentism (1920–1930)

Karl PEARSON (1857–1936)



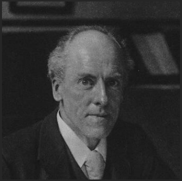
- He was a social darwinist & eugenist.
- The Grammar of Science (1892) influenced Einstein.
- Developed: (1) χ^2 test; (2) standard-deviation; (3) correlation; (4) p-value; (5) P.C.A.
- Initiator of the anti-Bayesian current.

Sir Ronald FISHER (1890–1962)

- Director of the department of eugenics at UCL.

The Golden Age of Frequentism (1920–1930)

Karl PEARSON (1857–1936)



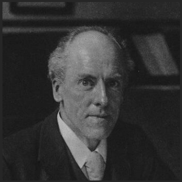
- He was a social darwinist & eugenist.
- The Grammar of Science (1892) influenced Einstein.
- Developed: (1) χ^2 test; (2) standard-deviation; (3) correlation; (4) p-value; (5) P.C.A.
- Initiator of the anti-Bayesian current.

Sir Ronald FISHER (1890–1962)

- Director of the department of eugenics at UCL.
- Developed: (1) maximum likelihood; (2) F-distribution/F-test; (3) null hypothesis & $p < 0.05$ (Statistical Methods for Research Workers, 1925).

The Golden Age of Frequentism (1920–1930)

Karl PEARSON (1857–1936)

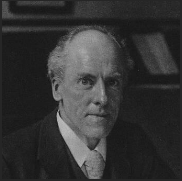


- He was a social darwinist & eugenist.
- The Grammar of Science (1892) influenced Einstein.
- Developed: (1) χ^2 test; (2) standard-deviation; (3) correlation; (4) p-value; (5) P.C.A.
- Initiator of the anti-Bayesian current.

Sir Ronald FISHER (1890–1962)

- Director of the department of eugenics at UCL.
- Developed: (1) maximum likelihood; (2) F-distribution/F-test; (3) null hypothesis & $p < 0.05$ (Statistical Methods for Research Workers, 1925).
- Paid as a consultant by the *Tobacco Manufacturers' Standing Committee* ⇒ publicly against the 1950 study showing that tobacco causes lung cancer (*"correlation does not imply causation"*; British Medical Journal, 1957).

Karl PEARSON (1857–1936)



- He was a social darwinist & eugenist.
- The Grammar of Science (1892) influenced Einstein.
- Developed: (1) χ^2 test; (2) standard-deviation; (3) correlation; (4) p-value; (5) P.C.A.
- Initiator of the anti-Bayesian current.

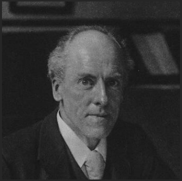
Jerzy NEYMAN (1894–1981)

Sir Ronald FISHER (1890–1962)

- Director of the department of eugenics at UCL.
- Developed: (1) maximum likelihood; (2) F-distribution/F-test; (3) null hypothesis & $p < 0.05$ (Statistical Methods for Research Workers, 1925).
- Paid as a consultant by the *Tobacco Manufacturers' Standing Committee* ⇒ publicly against the 1950 study showing that tobacco causes lung cancer (*"correlation does not imply causation"*; British Medical Journal, 1957).

The Golden Age of Frequentism (1920–1930)

Karl PEARSON (1857–1936)



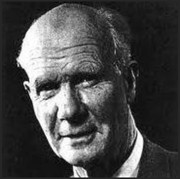
- He was a social darwinist & eugenist.
- The Grammar of Science (1892) influenced Einstein.
- Developed: (1) χ^2 test; (2) standard-deviation; (3) correlation; (4) p-value; (5) P.C.A.
- Initiator of the anti-Bayesian current.

Sir Ronald FISHER (1890–1962)

- Director of the department of eugenics at UCL.
- Developed: (1) maximum likelihood; (2) F-distribution/F-test; (3) null hypothesis & $p < 0.05$ (Statistical Methods for Research Workers, 1925).
- Paid as a consultant by the *Tobacco Manufacturers' Standing Committee* ⇒ publicly against the 1950 study showing that tobacco causes lung cancer (*"correlation does not imply causation"*; British Medical Journal, 1957).

Jerzy NEYMAN (1894–1981)

Egon PEARSON (1895–1980)



Bayes at Bletchley Park

Alan TURING (1912–1954)

Bayes at Bletchley Park

Alan TURING (1912–1954)

- Founder of theoretical computer science & artificial intelligence.

Bayes at Bletchley Park

Alan TURING (1912–1954) 🇬🇧

- Founder of theoretical computer science & artificial intelligence.
- Formalised the concepts of algorithm, computability & Turing machine.

Alan TURING (1912–1954)

- Founder of theoretical computer science & artificial intelligence.
- Formalised the concepts of algorithm, computability & Turing machine.

The Enigma Code Breaking:

Bayes at Bletchley Park

Alan TURING (1912–1954)

- Founder of theoretical computer science & artificial intelligence.
- Formalised the concepts of algorithm, computability & Turing machine.

The Enigma Code Breaking:

German submarines (U-boats) coded their communication with cryptographic *Enigma* machines.

Alan TURING (1912–1954)

- Founder of theoretical computer science & artificial intelligence.
- Formalised the concepts of algorithm, computability & Turing machine.

The Enigma Code Breaking:

German submarines (U-boats) coded their communication with cryptographic *Enigma* machines.

- Turing built a mechanical computer (*The Bomb*) which tested the combinations.

Alan TURING (1912–1954) 🇬🇧

- Founder of theoretical computer science & artificial intelligence.
- Formalised the concepts of algorithm, computability & Turing machine.

The Enigma Code Breaking:

German submarines (U-boats) coded their communication with cryptographic *Enigma* machines.

- Turing built a mechanical computer (*The Bomb*) which tested the combinations.
- He used *Bayesian priors* to reduce the number of combinations, looking for frequent words, such as *"ein"*, *"Heil Hitler"* & meteorological terms.

Bayes at Bletchley Park

Alan TURING (1912–1954)

- Founder of theoretical computer science & artificial intelligence.
- Formalised the concepts of algorithm, computability & Turing machine.

The Enigma Code Breaking:

German submarines (U-boats) coded their communication with cryptographic *Enigma* machines.

- Turing built a mechanical computer (*The Bomb*) which tested the combinations.
- He used *Bayesian priors* to reduce the number of combinations, looking for frequent words, such as *"ein"*, *"Heil Hitler"* & meteorological terms.
- He developed a unit quantifying the weight of evidence (Bayes factor), named the *ban*, after the city of Banbury, where punch cards were printed.

Bayes at Bletchley Park

Alan TURING (1912–1954) 

- Founder of theoretical computer science & artificial intelligence.
- Formalised the concepts of algorithm, computability & Turing machine.

The Enigma Code Breaking:

German submarines (U-boats) coded their communication with cryptographic *Enigma* machines.

- Turing built a mechanical computer (*The Bomb*) which tested the combinations.
- He used *Bayesian priors* to reduce the number of combinations, looking for frequent words, such as *"ein"*, *"Heil Hitler"* & meteorological terms.
- He developed a unit quantifying the weight of evidence (Bayes factor), named the *ban*, after the city of Banbury, where punch cards were printed.

The End of the War:

- In 1944, Thomas FLOWERS rused vacuum tubes instead of mechanical parts $\Rightarrow$ *The Colossus* computer.

Alan TURING (1912–1954)

- Founder of theoretical computer science & artificial intelligence.
- Formalised the concepts of algorithm, computability & Turing machine.

The Enigma Code Breaking:

German submarines (U-boats) coded their communication with cryptographic *Enigma* machines.

- Turing built a mechanical computer (*The Bomb*) which tested the combinations.
- He used *Bayesian priors* to reduce the number of combinations, looking for frequent words, such as *"ein"*, *"Heil Hitler"* & meteorological terms.
- He developed a unit quantifying the weight of evidence (Bayes factor), named the *ban*, after the city of Banbury, where punch cards were printed.

The End of the War:

- In 1944, Thomas FLOWERS rused vacuum tubes instead of mechanical parts $\Rightarrow$ *The Colossus* computer.
- Decodes Hitler's intentions in case of landing in Normandie $\Rightarrow$ shorten the war by two years (Eisenhower).

Alan TURING (1912–1954) 

- Founder of theoretical computer science & artificial intelligence.
- Formalised the concepts of algorithm, computability & Turing machine.

The Enigma Code Breaking:

German submarines (U-boats) coded their communication with cryptographic *Enigma* machines.

- Turing built a mechanical computer (*The Bomb*) which tested the combinations.
- He used *Bayesian priors* to reduce the number of combinations, looking for frequent words, such as *"ein"*, *"Heil Hitler"* & meteorological terms.
- He developed a unit quantifying the weight of evidence (Bayes factor), named the *ban*, after the city of Banbury, where punch cards were printed.

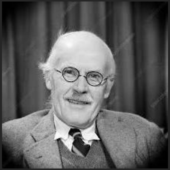
The End of the War:

- In 1944, Thomas FLOWERS rused vacuum tubes instead of mechanical parts $\Rightarrow$ *The Colossus* computer.
- Decodes Hitler's intentions in case of landing in Normandie $\Rightarrow$ shorten the war by two years (Eisenhower).
- The work at Bletchley Park remained *classified until 1973*, to hide that the British could crack Russian codes $\Rightarrow$ delayed glory for the Bayesian approach.

The Bayesian Resistance

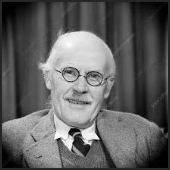
The Bayesian Resistance

Sir Harold JEFFREYS (1891–1989)



The Bayesian Resistance

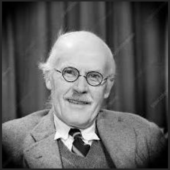
Sir Harold JEFFREYS (1891–1989)



- Geophysicist & mathematician (University of Cambridge).

The Bayesian Resistance

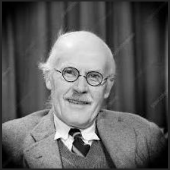
Sir Harold JEFFREYS (1891–1989)



- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) $\Rightarrow$ liquid Earth core.

The Bayesian Resistance

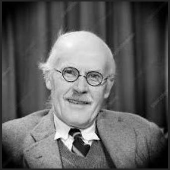
Sir Harold JEFFREYS (1891–1989)



- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) ⇒ liquid Earth core.
- Initiated the Bayesian revival (Theory of Probability, 1939).

The Bayesian Resistance

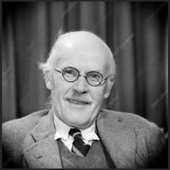
Sir Harold JEFFREYS (1891–1989)



- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) $\Rightarrow$ liquid Earth core.
- Initiated the Bayesian revival (Theory of Probability, 1939).
- Popularized the use of Bayes factors.

The Bayesian Resistance

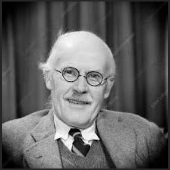
Sir Harold JEFFREYS (1891–1989)



- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) $\Rightarrow$ liquid Earth core.
- Initiated the Bayesian revival (Theory of Probability, 1939).
- Popularized the use of Bayes factors.
- Opposed to continental drift, though.

The Bayesian Resistance

Sir Harold JEFFREYS (1891–1989)

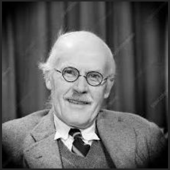


- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) $\Rightarrow$ liquid Earth core.
- Initiated the Bayesian revival (Theory of Probability, 1939).
- Popularized the use of Bayes factors.
- Opposed to continental drift, though.

Post World War II Era:

The Bayesian Resistance

Sir Harold JEFFREYS (1891–1989)



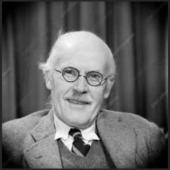
- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) ⇒ liquid Earth core.
- Initiated the Bayesian revival (Theory of Probability, 1939).
- Popularized the use of Bayes factors.
- Opposed to continental drift, though.

Post World War II Era:

Arthur BAILEY applied Bayes' rule, including the probability of events that had never happened ⇒ estimate insurance premiums.

The Bayesian Resistance

Sir Harold JEFFREYS (1891–1989)



- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) ⇒ liquid Earth core.
- Initiated the Bayesian revival (Theory of Probability, 1939).
- Popularized the use of Bayes factors.
- Opposed to continental drift, though.

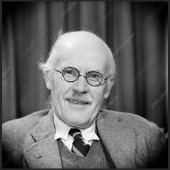
Post World War II Era:

Arthur BAILEY applied Bayes' rule, including the probability of events that had never happened ⇒ estimate insurance premiums.

Dennis LINDLEY & Jimmy SAVAGE popularised Bayes' rule (Savage: *"Fisher is making Bayesian omelet without breaking Bayesian eggs"*).

The Bayesian Resistance

Sir Harold JEFFREYS (1891–1989)



- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) ⇒ liquid Earth core.
- Initiated the Bayesian revival (Theory of Probability, 1939).
- Popularized the use of Bayes factors.
- Opposed to continental drift, though.

Post World War II Era:

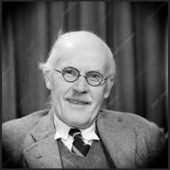
Arthur BAILEY applied Bayes' rule, including the probability of events that had never happened ⇒ estimate insurance premiums.

Dennis LINDLEY & Jimmy SAVAGE popularised Bayes' rule (Savage: *"Fisher is making Bayesian omelet without breaking Bayesian eggs"*).

Jerome CORNFIELD pioneered in applying Bayes' rule to epidemiology ⇒ showed link between smoking & lung cancer (ridiculised Fisher).

The Bayesian Resistance

Sir Harold JEFFREYS (1891–1989) 🇬🇧



- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) ⇒ liquid Earth core.
- Initiated the Bayesian revival (Theory of Probability, 1939).
- Popularized the use of Bayes factors.
- Opposed to continental drift, though.

Post World War II Era:

Arthur BAILEY 🇺🇸 applied Bayes' rule, including the probability of events that had never happened ⇒ estimate insurance premiums.

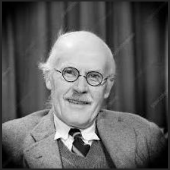
Dennis LINDLEY 🇬🇧 **& Jimmy SAVAGE** 🇺🇸 popularised Bayes' rule (Savage: *"Fisher is making Bayesian omelet without breaking Bayesian eggs"*).

Jerome CORNFIELD 🇺🇸 pioneered in applying Bayes' rule to epidemiology ⇒ showed link between smoking & lung cancer (ridiculised Fisher).

Howard RAIFFA & Robert SCHLAIFER 🇺🇸 taught Bayes' rule for business & decision-making.

The Bayesian Resistance

Sir Harold JEFFREYS (1891–1989)



- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) ⇒ liquid Earth core.
- Initiated the Bayesian revival (Theory of Probability, 1939).
- Popularized the use of Bayes factors.
- Opposed to continental drift, though.

Post World War II Era:

Arthur BAILEY applied Bayes' rule, including the probability of events that had never happened ⇒ estimate insurance premiums.

Dennis LINDLEY & Jimmy SAVAGE popularised Bayes' rule (Savage: *"Fisher is making Bayesian omelet without breaking Bayesian eggs"*).

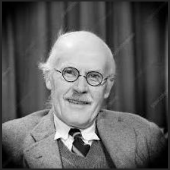
Jerome CORNFIELD pioneered in applying Bayes' rule to epidemiology ⇒ showed link between smoking & lung cancer (ridiculised Fisher).

Howard RAIFFA & Robert SCHLAIFER taught Bayes' rule for business & decision-making.

Norman RASMUSSEN estimated the risks of a nuclear incident in the 70s ⇒ possible, but not necessarily desastrous (*cf.* Three Mile Island, 1979).

The Bayesian Resistance

Sir Harold JEFFREYS (1891–1989)



- Geophysicist & mathematician (University of Cambridge).
- In 1926, Bayesian analysis of earthquakes (few data) ⇒ liquid Earth core.
- Initiated the Bayesian revival (Theory of Probability, 1939).
- Popularized the use of Bayes factors.
- Opposed to continental drift, though.

Post World War II Era:

Arthur BAILEY applied Bayes' rule, including the probability of events that had never happened ⇒ estimate insurance premiums.

Dennis LINDLEY & Jimmy SAVAGE popularised Bayes' rule (Savage: *"Fisher is making Bayesian omelet without breaking Bayesian eggs"*).

Jerome CORNFIELD pioneered in applying Bayes' rule to epidemiology ⇒ showed link between smoking & lung cancer (ridiculised Fisher).

Howard RAIFFA & Robert SCHLAIFER taught Bayes' rule for business & decision-making.

Norman RASMUSSEN estimated the risks of a nuclear incident in the 70s ⇒ possible, but not necessarily desastrous (*cf.* Three Mile Island, 1979).

Bayesian Search Algorithm used to find lost nuclear bombs & russian submarines.

The Great Numerical Leap Forward

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

1984: Most popular MCMC solver, the *Gibbs sampling* (Geman & Geman, 1984).

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

1984: Most popular MCMC solver, the *Gibbs sampling* (Geman & Geman, 1984).

$\Rightarrow$ Bayesian techniques became more attractive.

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

1984: Most popular MCMC solver, the *Gibbs sampling* (Geman & Geman, 1984).

⇒ Bayesian techniques became more attractive.

Modern Applications:

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

1984: Most popular MCMC solver, the *Gibbs sampling* (Geman & Geman, 1984).

$\Rightarrow$ Bayesian techniques became more attractive.

Modern Applications:

1983: NASA estimated the probability of shuttle failure at *1 in 100 000*, with frequentist methods. An independent Bayesian analysis estimated the odds of rocket booster failure at *1 in 35*. In 1986, the 28^{th} launch (Challenger) exploded.

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

1984: Most popular MCMC solver, the *Gibbs sampling* (Geman & Geman, 1984).

$\Rightarrow$ Bayesian techniques became more attractive.

Modern Applications:

1983: NASA estimated the probability of shuttle failure at *1 in 100 000*, with frequentist methods. An independent Bayesian analysis estimated the odds of rocket booster failure at *1 in 35*. In 1986, the 28th launch (Challenger) exploded.

1987: explosion of SN1987A, with two dozen neutrinos detected $\Rightarrow$ Loredo & Lamb (1989) sucessfully applied Bayesian modelling, while frequentist techniques were failing to analyze this valuable data (Loredo, 1990, for a review):

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

1984: Most popular MCMC solver, the *Gibbs sampling* (Geman & Geman, 1984).

$\Rightarrow$ Bayesian techniques became more attractive.

Modern Applications:

1983: NASA estimated the probability of shuttle failure at *1 in 100 000*, with frequentist methods. An independent Bayesian analysis estimated the odds of rocket booster failure at *1 in 35*. In 1986, the 28$^{\text{th}}$ launch (Challenger) exploded.

1987: explosion of SN1987A, with two dozen neutrinos detected $\Rightarrow$ Loredo & Lamb (1989) sucessfully applied Bayesian modelling, while frequentist techniques were failing to analyze this valuable data (Loredo, 1990, for a review): (1) very good agreement with theory of stellar collapse & neutron star formation; (2) upper limit on the mass of $\bar{\nu}_e$.

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

1984: Most popular MCMC solver, the *Gibbs sampling* (Geman & Geman, 1984).

$\Rightarrow$ Bayesian techniques became more attractive.

Modern Applications:

1983: NASA estimated the probability of shuttle failure at *1 in 100 000*, with frequentist methods. An independent Bayesian analysis estimated the odds of rocket booster failure at *1 in 35*. In 1986, the 28th launch (Challenger) exploded.

1987: explosion of SN1987A, with two dozen neutrinos detected $\Rightarrow$ Loredo & Lamb (1989) sucessfully applied Bayesian modelling, while frequentist techniques were failing to analyze this valuable data (Loredo, 1990, for a review): (1) very good agreement with theory of stellar collapse & neutron star formation; (2) upper limit on the mass of $\bar{\nu}_e$.

2003: completion of the *Human Genome Project*, which used Bayesian techniques.

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

1984: Most popular MCMC solver, the *Gibbs sampling* (Geman & Geman, 1984).

$\Rightarrow$ Bayesian techniques became more attractive.

Modern Applications:

1983: NASA estimated the probability of shuttle failure at *1 in 100 000*, with frequentist methods. An independent Bayesian analysis estimated the odds of rocket booster failure at *1 in 35*. In 1986, the 28^{th} launch (Challenger) exploded.

1987: explosion of SN1987A, with two dozen neutrinos detected $\Rightarrow$ Loredo & Lamb (1989) sucessfully applied Bayesian modelling, while frequentist techniques were failing to analyze this valuable data (Loredo, 1990, for a review): (1) very good agreement with theory of stellar collapse & neutron star formation; (2) upper limit on the mass of $\bar{\nu}_e$.

2003: completion of the *Human Genome Project*, which used Bayesian techniques.

The Bayesian Foundation of Machine-Learning & A.I.:

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

1984: Most popular MCMC solver, the *Gibbs sampling* (Geman & Geman, 1984).

$\Rightarrow$ Bayesian techniques became more attractive.

Modern Applications:

1983: NASA estimated the probability of shuttle failure at *1 in 100 000*, with frequentist methods. An independent Bayesian analysis estimated the odds of rocket booster failure at *1 in 35*. In 1986, the 28th launch (Challenger) exploded.

1987: explosion of SN1987A, with two dozen neutrinos detected $\Rightarrow$ Loredo & Lamb (1989) sucessfully applied Bayesian modelling, while frequentist techniques were failing to analyze this valuable data (Loredo, 1990, for a review): (1) very good agreement with theory of stellar collapse & neutron star formation; (2) upper limit on the mass of $\bar{\nu}_e$.

2003: completion of the *Human Genome Project*, which used Bayesian techniques.

The Bayesian Foundation of Machine-Learning & A.I.:

- Most machine-learning techniques are probabilistic.

The Great Numerical Leap Forward

Numerical Techniques to solve Bayesian Problems:

1970: First Markov Chain Monte-Carlo (MCMC) method, the *Metropolis-Hastings* algorithm (Hastings, 1970).

1984: Most popular MCMC solver, the *Gibbs sampling* (Geman & Geman, 1984).

$\Rightarrow$ Bayesian techniques became more attractive.

Modern Applications:

1983: NASA estimated the probability of shuttle failure at *1 in 100 000*, with frequentist methods. An independent Bayesian analysis estimated the odds of rocket booster failure at *1 in 35*. In 1986, the 28^{th} launch (Challenger) exploded.

1987: explosion of SN1987A, with two dozen neutrinos detected $\Rightarrow$ Loredo & Lamb (1989) sucessfully applied Bayesian modelling, while frequentist techniques were failing to analyze this valuable data (Loredo, 1990, for a review): (1) very good agreement with theory of stellar collapse & neutron star formation; (2) upper limit on the mass of $\bar{\nu}_e$.

2003: completion of the *Human Genome Project*, which used Bayesian techniques.

The Bayesian Foundation of Machine-Learning & A.I.:

- Most machine-learning techniques are probabilistic.
- Training a neural network $\Leftrightarrow$ informing a *prior*.

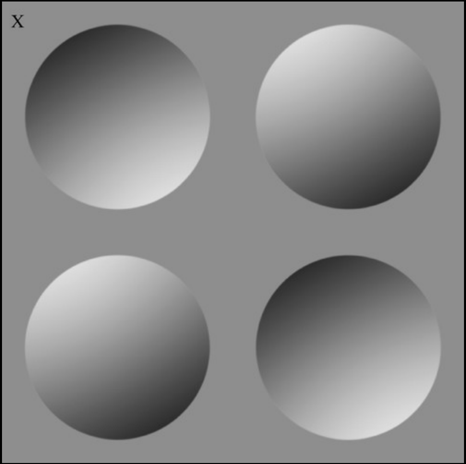
Importance of Bayes' Rule For Neurosciences

Importance of Bayes' Rule For Neurosciences

(Lectures given between 2011 and 2012)

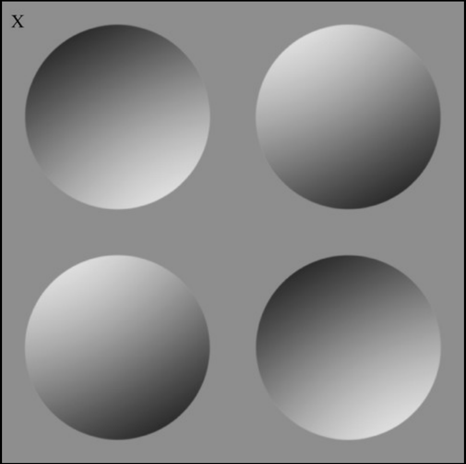
Optical Illusions: Manifestation of the Brain's Bayesian Prior

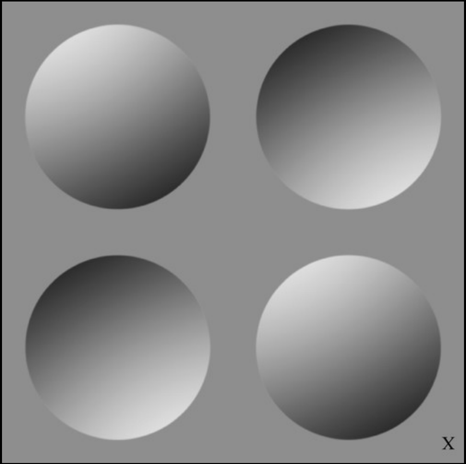
Optical Illusions: Manifestation of the Brain's Bayesian Prior



(From Dehaene's Lecture in College de France, 2011)

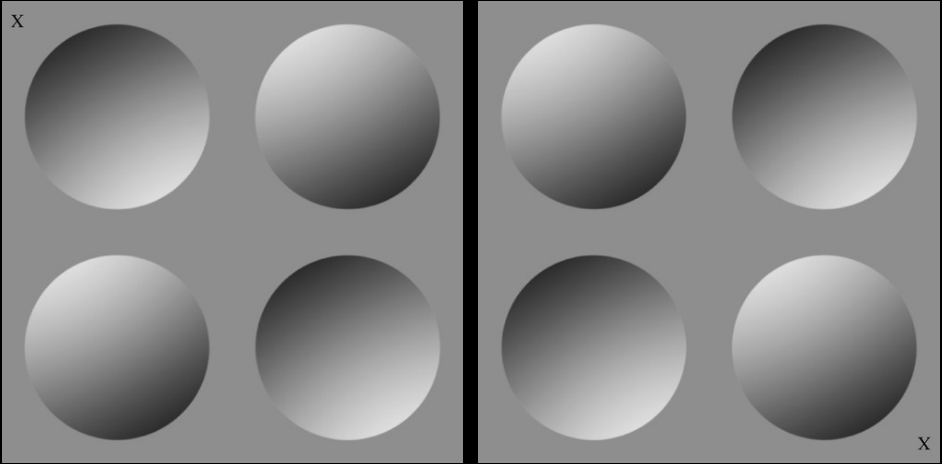
Optical Illusions: Manifestation of the Brain's Bayesian Prior





(From Dehaene's Lecture in College de France, 2011)

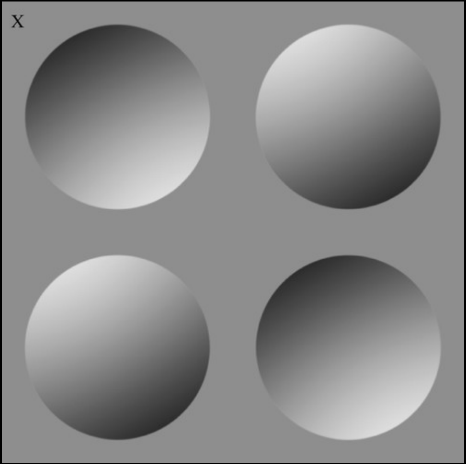
Optical Illusions: Manifestation of the Brain's Bayesian Prior

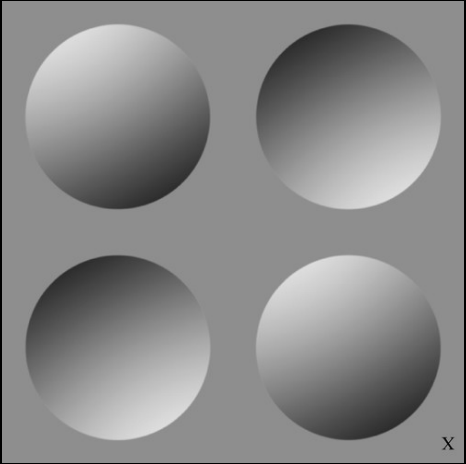


(From Dehaene's Lecture in College de France, 2011)

Most objects are illuminated from above (sunlight, spots, *etc.*)

Optical Illusions: Manifestation of the Brain's Bayesian Prior



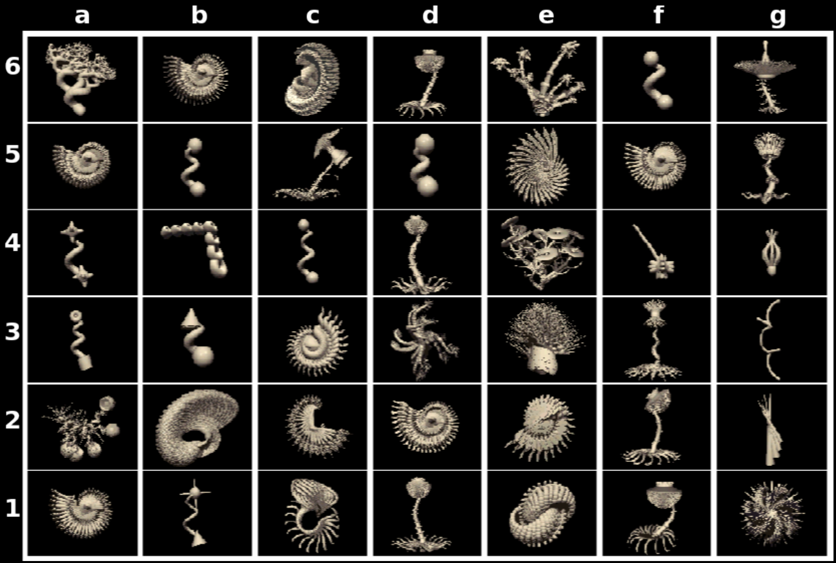


(From Dehaene's Lecture in College de France, 2011)

Most objects are illuminated from above (sunlight, spots, *etc.*) $\Rightarrow$ the visual cortex interprets shades, using this prior.

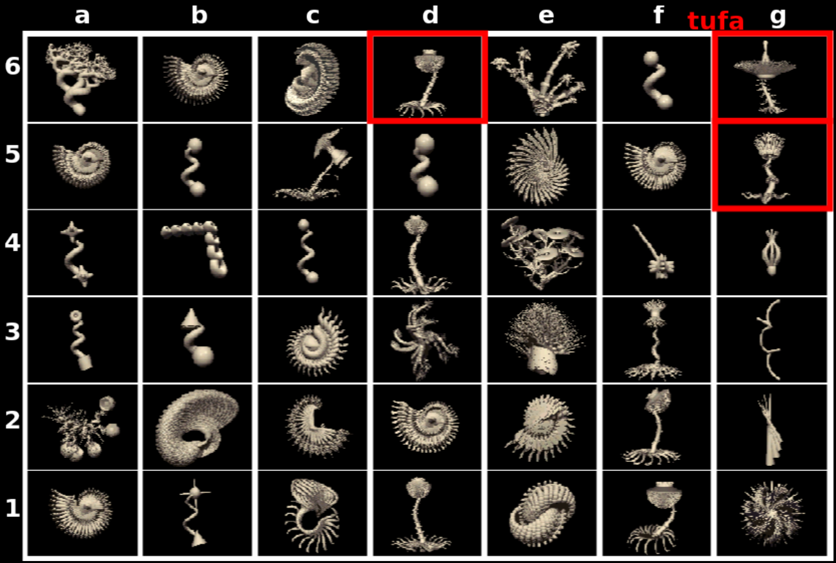
How Does the Brain Learn New Words?

How Does the Brain Learn New Words?



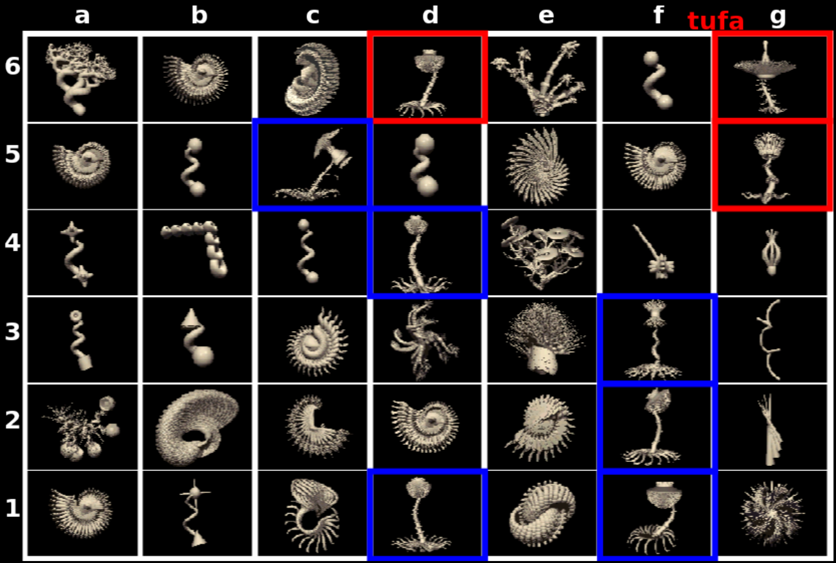
(Tenenbaum *et al.*, 2011)

How Does the Brain Learn New Words?



(Tenenbaum *et al.*, 2011)

How Does the Brain Learn New Words?



(Tenenbaum *et al.*, 2011)

Outline of the Talk

The Epistemological Debate at the Beginning of the XX$^{\text{th}}$ Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

The Epistemological Debate at the Beginning of the XX^{th} Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

Preliminary Definitions:

The Epistemological Debate at the Beginning of the XX$^{\text{th}}$ Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

Preliminary Definitions:

Epistemology: the philosophical study of the nature, origin, and limits of human knowledge.

The Epistemological Debate at the Beginning of the XXth Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

Preliminary Definitions:

Epistemology: the philosophical study of the nature, origin, and limits of human knowledge.

Deduction: inferring the truth of a specific case from general rules.

The Epistemological Debate at the Beginning of the XX$^{\text{th}}$ Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

Preliminary Definitions:

Epistemology: the philosophical study of the nature, origin, and limits of human knowledge.

Deduction: inferring the truth of a specific case from general rules.

Induction: inferring a general conclusion based on individual cases.

The Epistemological Debate at the Beginning of the XX$^{\text{th}}$ Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

Preliminary Definitions:

Epistemology: the philosophical study of the nature, origin, and limits of human knowledge.

Deduction: inferring the truth of a specific case from general rules.

Induction: inferring a general conclusion based on individual cases.

Scientific Positivism:

Auguste COMTE (1798–1857) 

The Epistemological Debate at the Beginning of the XX$^{\text{th}}$ Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

Preliminary Definitions:

Epistemology: the philosophical study of the nature, origin, and limits of human knowledge.

Deduction: inferring the truth of a specific case from general rules.

Induction: inferring a general conclusion based on individual cases.

Scientific Positivism:

Auguste COMTE (1798–1857)

- Aim at demarcating itself from theology & metaphysics.

The Epistemological Debate at the Beginning of the XX$^{\text{th}}$ Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

Preliminary Definitions:

Epistemology: the philosophical study of the nature, origin, and limits of human knowledge.

Deduction: inferring the truth of a specific case from general rules.

Induction: inferring a general conclusion based on individual cases.

Scientific Positivism:

Auguste COMTE (1798–1857)

- Aim at demarcating itself from theology & metaphysics.
- Renounce to understand the *causes* (why), to focus on the mathematical *laws* of nature (how).

The Epistemological Debate at the Beginning of the XX$^{\text{th}}$ Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

Preliminary Definitions:

Epistemology: the philosophical study of the nature, origin, and limits of human knowledge.

Deduction: inferring the truth of a specific case from general rules.

Induction: inferring a general conclusion based on individual cases.

Scientific Positivism:

Auguste COMTE (1798–1857)

- Aim at demarcating itself from theology & metaphysics.
- Renounce to understand the *causes* (why), to focus on the mathematical *laws* of nature (how).

Conventionalism & Verificationism:

Henri POINCARÉ (1854–1912)

The Epistemological Debate at the Beginning of the XXth Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

Preliminary Definitions:

Epistemology: the philosophical study of the nature, origin, and limits of human knowledge.

Deduction: inferring the truth of a specific case from general rules.

Induction: inferring a general conclusion based on individual cases.

Scientific Positivism:

Auguste COMTE (1798–1857) 🇫🇷

- Aim at demarcating itself from theology & metaphysics.
- Renounce to understand the *causes* (why), to focus on the mathematical *laws* of nature (how).

Conventionalism & Verificationism:

Henri POINCARÉ (1854–1912) 🇫🇷

- *Human intuitions* about the physical world are possibly flawed (*e.g.* Euclidian geometry).

The Epistemological Debate at the Beginning of the XX$^{\text{th}}$ Century

The unity of science consists alone in its method, not in its material.

Karl PEARSON (The Grammar of Science, 1892)

Preliminary Definitions:

Epistemology: the philosophical study of the nature, origin, and limits of human knowledge.

Deduction: inferring the truth of a specific case from general rules.

Induction: inferring a general conclusion based on individual cases.

Scientific Positivism:

Auguste COMTE (1798–1857)

- Aim at demarcating itself from theology & metaphysics.
- Renounce to understand the *causes* (why), to focus on the mathematical *laws* of nature (how).

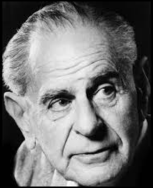
Conventionalism & Verificationism:

Henri POINCARÉ (1854–1912)

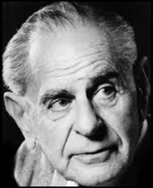
- *Human intuitions* about the physical world are possibly flawed (*e.g.* Euclidian geometry).
- Abandon rationalism for empiricism $\Rightarrow$ a proposition has a cognitive meaning only if it can be *verified* by experience.

Popper's Criticism of Induction in Empirical Sciences

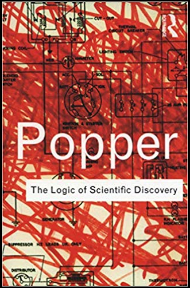
Popper's Criticism of Induction in Empirical Sciences



Karl POPPER
(1902–1994)

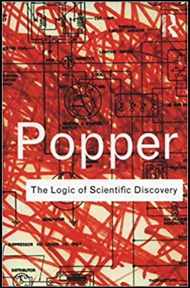


Karl POPPER
(1902–1994)



Very influential book, still a reference today (published in 1935; English version in 1959).

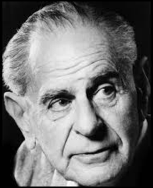
Karl POPPER
(1902–1994)



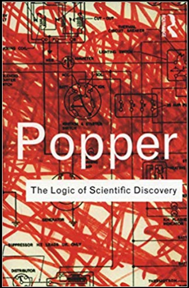
Very influential book, still a reference today (published in 1935; English version in 1959).

The Myth of Induction:

Popper's Criticism of Induction in Empirical Sciences



Karl POPPER
(1902–1994)



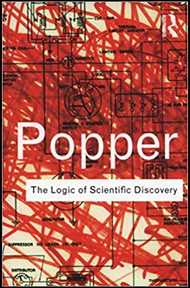
Very influential book, still a reference today (published in 1935; English version in 1959).

The Myth of Induction:

- Inductive logic *"does not provide a suitable criterion of demarcation with metaphysical speculation"*.

Popper's Criticism of Induction in Empirical Sciences

Karl POPPER
(1902–1994)

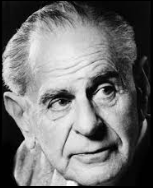


Very influential book, still a reference today (published in 1935; English version in 1959).

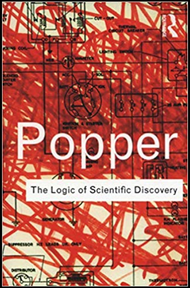
The Myth of Induction:

- Inductive logic *"does not provide a suitable criterion of demarcation with metaphysical speculation"*.
- *"The actual procedure of science is to operate with conjectures: to jump to conclusions – often after one single observation"*.

Popper's Criticism of Induction in Empirical Sciences



Karl POPPER
(1902–1994)



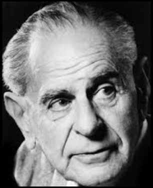
Very influential book, still a reference today (published in 1935; English version in 1959).

The Myth of Induction:

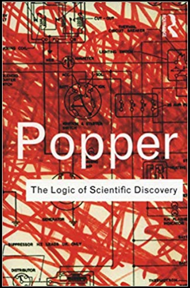
- Inductive logic *"does not provide a suitable criterion of demarcation with metaphysical speculation"*.
- *"The actual procedure of science is to operate with conjectures: to jump to conclusions – often after one single observation"*.

$\Rightarrow$ Deductivism: *"Hypotheses can only be emprically tested and only after they have been advanced"*.

Popper's Criticism of Induction in Empirical Sciences



Karl POPPER
(1902–1994)



Very influential book, still a reference today (published in 1935; English version in 1959).

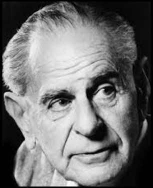
The Myth of Induction:

- Inductive logic *"does not provide a suitable criterion of demarcation with metaphysical speculation"*.
- *"The actual procedure of science is to operate with conjectures: to jump to conclusions – often after one single observation"*.

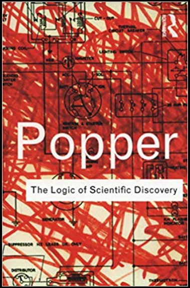
⇒ Deductivism: *"Hypotheses can only be emprically tested and only after they have been advanced"*.

Falsifiability, the Criterion of Demarcation:

Popper's Criticism of Induction in Empirical Sciences



Karl POPPER
(1902–1994)



Very influential book, still a reference today (published in 1935; English version in 1959).

The Myth of Induction:

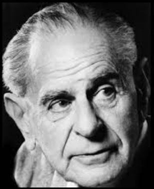
- Inductive logic *"does not provide a suitable criterion of demarcation with metaphysical speculation"*.
- *"The actual procedure of science is to operate with conjectures: to jump to conclusions – often after one single observation"*.

$\Rightarrow$ Deductivism: *"Hypotheses can only be emprically tested and only after they have been advanced"*.

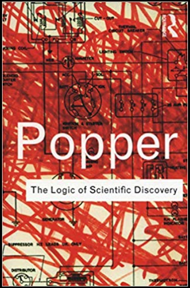
Falsifiability, the Criterion of Demarcation:

- Criticize conventionalism who *"evade falsification by using ad hoc modifications of the theory."*

Popper's Criticism of Induction in Empirical Sciences



Karl POPPER
(1902–1994)



Very influential book, still a reference today (published in 1935; English version in 1959).

The Myth of Induction:

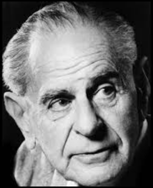
- Inductive logic *"does not provide a suitable criterion of demarcation with metaphysical speculation"*.
- *"The actual procedure of science is to operate with conjectures: to jump to conclusions – often after one single observation"*.

$\Rightarrow$ Deductivism: *"Hypotheses can only be emprically tested and only after they have been advanced"*.

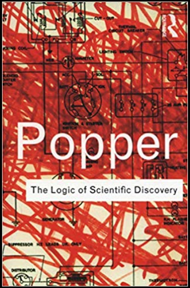
Falsifiability, the Criterion of Demarcation:

- Criticize conventionalism who *"evade falsification by using ad hoc modifications of the theory."*
- *"One must not save from falsification a theory if it has failed."*

Popper's Criticism of Induction in Empirical Sciences



Karl POPPER
(1902–1994)



Very influential book, still a reference today (published in 1935; English version in 1959).

The Myth of Induction:

- Inductive logic *"does not provide a suitable criterion of demarcation with metaphysical speculation"*.
- *"The actual procedure of science is to operate with conjectures: to jump to conclusions – often after one single observation"*.

$\Rightarrow$ Deductivism: *"Hypotheses can only be emprically tested and only after they have been advanced"*.

Falsifiability, the Criterion of Demarcation:

- Criticize conventionalism who *"evade falsification by using ad hoc modifications of the theory."*
- *"One must not save from falsification a theory if it has failed."*

$\Rightarrow$ Falsifiabilism: deductivism & *modus tollens*.
Modus tollens: $\big((A \Rightarrow B) \wedge \overline{B}\big) \Rightarrow \overline{A}$.

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone $\Rightarrow$ no coincidence.

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone ⇒ no coincidence.

Principle of Parsimony:

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone $\Rightarrow$ no coincidence.

Principle of Parsimony:

Simplicity $\Leftrightarrow$ degree of falsifiability $\Leftrightarrow$ higher prior improbability.

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone $\Rightarrow$ no coincidence.

Principle of Parsimony:

Simplicity $\Leftrightarrow$ degree of falsifiability $\Leftrightarrow$ higher prior improbability.
$\Rightarrow$ simpler theories have a higher *empirical content*.

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone $\Rightarrow$ no coincidence.

Principle of Parsimony:

Simplicity $\Leftrightarrow$ degree of falsifiability $\Leftrightarrow$ higher prior improbability.
$\Rightarrow$ simpler theories have a higher *empirical content*.

Popper's Epistemology has a Frequentist Frame of Mind:

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone $\Rightarrow$ no coincidence.

Principle of Parsimony:

Simplicity $\Leftrightarrow$ degree of falsifiability $\Leftrightarrow$ higher prior improbability.
$\Rightarrow$ simpler theories have a higher *empirical content*.

Popper's Epistemology has a Frequentist Frame of Mind:

Falsifiability $\Leftrightarrow$ p-value to reject the null hypothesis.

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone $\Rightarrow$ no coincidence.

Principle of Parsimony:

Simplicity $\Leftrightarrow$ degree of falsifiability $\Leftrightarrow$ higher prior improbability.
$\Rightarrow$ simpler theories have a higher *empirical content*.

Popper's Epistemology has a Frequentist Frame of Mind:

Falsifiability $\Leftrightarrow$ p-value to reject the null hypothesis.

Boolean/Platonic logic & reject of probability as quantification of knowledge (no account for its subjectivity).

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone $\Rightarrow$ no coincidence.

Principle of Parsimony:

Simplicity $\Leftrightarrow$ degree of falsifiability $\Leftrightarrow$ higher prior improbability.
$\Rightarrow$ simpler theories have a higher *empirical content*.

Popper's Epistemology has a Frequentist Frame of Mind:

Falsifiability $\Leftrightarrow$ p-value to reject the null hypothesis.

Boolean/Platonic logic & reject of probability as quantification of knowledge (no account for its subjectivity).

Necessary repeatability: no possibility to account for sparse, unique constraints (must first think about a falsifiable experience).

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone ⇒ no coincidence.

Principle of Parsimony:

Simplicity ⇔ degree of falsifiability ⇔ higher prior improbability.
⇒ simpler theories have a higher *empirical content*.

Popper's Epistemology has a Frequentist Frame of Mind:

Falsifiability ⇔ *p*-value to reject the null hypothesis.

Boolean/Platonic logic & reject of probability as quantification of knowledge (no account for its subjectivity).

Necessary repeatability: no possibility to account for sparse, unique constraints (must first think about a falsifiable experience).

Parsimony: *ad hoc* principle.

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone $\Rightarrow$ no coincidence.

Principle of Parsimony:

Simplicity $\Leftrightarrow$ degree of falsifiability $\Leftrightarrow$ higher prior improbability.
$\Rightarrow$ simpler theories have a higher *empirical content*.

Popper's Epistemology has a Frequentist Frame of Mind:

Falsifiability $\Leftrightarrow$ p-value to reject the null hypothesis.

Boolean/Platonic logic & reject of probability as quantification of knowledge (no account for its subjectivity).

Necessary repeatability: no possibility to account for sparse, unique constraints (must first think about a falsifiable experience).

Parsimony: *ad hoc* principle.

Opposed to induction: *"The logic of probable inference leads to apriorism"*.

Objectivity & Principle of Parsimony (Occam's Razor)

Objectivity & Reproducibility:

- The objectivity of scientific statements lie in the fact that they can be *inter-subjectively tested*.
- Only repeatable experiments can be tested by anyone $\Rightarrow$ no coincidence.

Principle of Parsimony:

Simplicity $\Leftrightarrow$ degree of falsifiability $\Leftrightarrow$ higher prior improbability.
$\Rightarrow$ simpler theories have a higher *empirical content*.

Popper's Epistemology has a Frequentist Frame of Mind:

Falsifiability $\Leftrightarrow$ p-value to reject the null hypothesis.

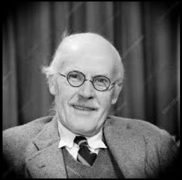
Boolean/Platonic logic & reject of probability as quantification of knowledge (no account for its subjectivity).

Necessary repeatability: no possibility to account for sparse, unique constraints (must first think about a falsifiable experience).

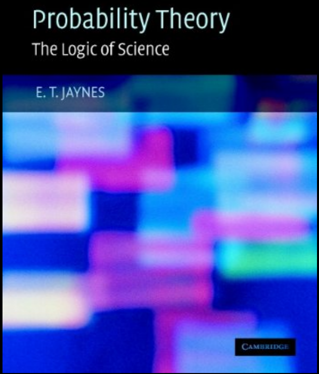
Parsimony: *ad hoc* principle.

Opposed to induction: *"The logic of probable inference leads to apriorism"*.

$\Rightarrow$ It was conceived at the peak of the *frequentist winter*.

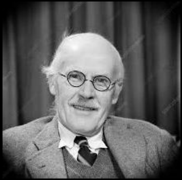


Sir Harold JEFFREYS
(1891–1989) 🇬🇧

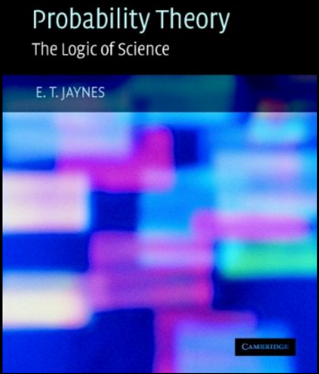


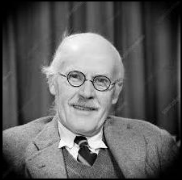
(Published in 2003; dedicated to Jeffreys)

Edwin Thompson JAYNES (1922–1998)



Sir Harold JEFFREYS (1891–1989)

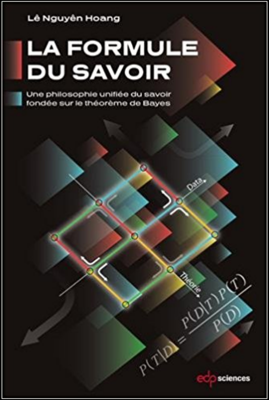




Sir Harold JEFFREYS (1891–1989)

(Published in 2003; dedicated to Jeffreys)

Edwin Thompson JAYNES (1922–1998)



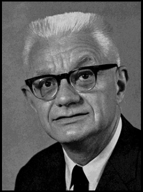
Lê Nguyên HOANG (1987–)

(Published in 2018)

Falsifiability & the Limits of Platonic Logic

Falsifiability & the Limits of Platonic Logic

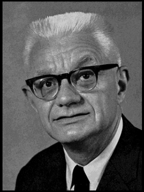
Hempel's Paradox (Hempel, 1940):



Carl Gustav
HEMPEL
(1905–1997)

Hempel's Paradox (Hempel, 1940):

$$\underbrace{\text{Raven} \Rightarrow \text{Black}}_{\textit{proposition}}$$

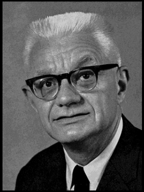


Carl Gustav HEMPEL (1905–1997)

Falsifiability & the Limits of Platonic Logic

Hempel's Paradox (Hempel, 1940):

$$\underbrace{\text{Raven} \Rightarrow \text{Black}}_{\textit{proposition}} \quad \Leftrightarrow \quad \underbrace{\text{Not Black} \Rightarrow \text{Not a Raven}}_{\textit{contraposition}}$$



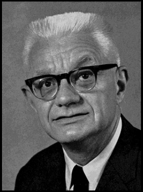
Carl Gustav HEMPEL (1905–1997)

Falsifiability & the Limits of Platonic Logic

Hempel's Paradox (Hempel, 1940):

$$\underbrace{\text{Raven} \Rightarrow \text{Black}}_{\textit{proposition}} \Leftrightarrow \underbrace{\text{Not Black} \Rightarrow \text{Not a Raven}}_{\textit{contraposition}}$$

Thus:  $\Longrightarrow$ 



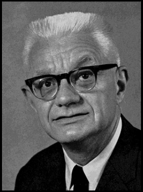
Carl Gustav HEMPEL (1905–1997)

Falsifiability & the Limits of Platonic Logic

Hempel's Paradox (Hempel, 1940):

$$\underbrace{\text{Raven} \Rightarrow \text{Black}}_{\textit{proposition}} \quad \Leftrightarrow \quad \underbrace{\text{Not Black} \Rightarrow \text{Not a Raven}}_{\textit{contraposition}}$$

Thus:  $\Longrightarrow$ 



Carl Gustav HEMPEL (1905–1997)

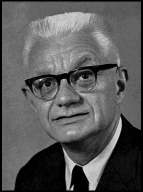
Bayesian Solution to the Paradox:

$$\text{Raven} \Rightarrow \text{Black} \quad \Leftrightarrow \quad p(\text{Black}|\text{Raven}) = 1$$

Hempel's Paradox (Hempel, 1940):

$$\underbrace{\text{Raven} \Rightarrow \text{Black}}_{\textit{proposition}} \quad \Leftrightarrow \quad \underbrace{\text{Not Black} \Rightarrow \text{Not a Raven}}_{\textit{contraposition}}$$

Thus:  $\Longrightarrow$ 



Carl Gustav HEMPEL (1905–1997)

Bayesian Solution to the Paradox:

$$\text{Raven} \Rightarrow \text{Black} \quad \Leftrightarrow \quad p(\text{Black}|\text{Raven}) = 1$$

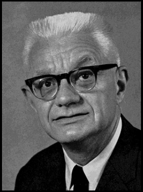
$$p(\text{Not a Raven}|\text{Not Black}) = 1 - \frac{1 - p(\text{Black}|\text{Raven})}{1 - p(\text{Black})} p(\text{Raven})$$

Falsifiability & the Limits of Platonic Logic

Hempel's Paradox (Hempel, 1940):

$$\underbrace{\text{Raven} \Rightarrow \text{Black}}_{\textit{proposition}} \quad \Leftrightarrow \quad \underbrace{\text{Not Black} \Rightarrow \text{Not a Raven}}_{\textit{contraposition}}$$

Thus:  $\Longrightarrow$ 



Carl Gustav HEMPEL (1905–1997)

Bayesian Solution to the Paradox:

$$\text{Raven} \Rightarrow \text{Black} \quad \Leftrightarrow \quad p(\text{Black}|\text{Raven}) = 1$$

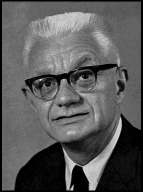
$$p(\text{Not a Raven}|\text{Not Black}) = 1 - \frac{1 - p(\text{Black}|\text{Raven})}{1 - p(\text{Black})} p(\text{Raven}) \neq p(\text{Black}|\text{Raven})$$

Falsifiability & the Limits of Platonic Logic

Hempel's Paradox (Hempel, 1940):

$$\underbrace{\text{Raven} \Rightarrow \text{Black}}_{\textit{proposition}} \quad \Leftrightarrow \quad \underbrace{\text{Not Black} \Rightarrow \text{Not a Raven}}_{\textit{contraposition}}$$

Thus:  $\Longrightarrow$ 



Carl Gustav HEMPEL (1905–1997)

Bayesian Solution to the Paradox:

$$\text{Raven} \Rightarrow \text{Black} \quad \Leftrightarrow \quad p(\text{Black}|\text{Raven}) = 1$$

$$p(\text{Not a Raven}|\text{Not Black}) = 1 - \frac{1 - p(\text{Black}|\text{Raven})}{1 - p(\text{Black})} p(\text{Raven}) \neq p(\text{Black}|\text{Raven})$$

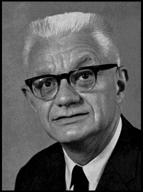
Most importantly: the weight of evidence (Bayes factor) is small for a red apple (Good, 1960).

Falsifiability & the Limits of Platonic Logic

Hempel's Paradox (Hempel, 1940):

$$\underbrace{\text{Raven} \Rightarrow \text{Black}}_{\textit{proposition}} \quad \Leftrightarrow \quad \underbrace{\text{Not Black} \Rightarrow \text{Not a Raven}}_{\textit{contraposition}}$$

Thus:  $\Longrightarrow$ 



Carl Gustav HEMPEL (1905–1997)

Bayesian Solution to the Paradox:

$$\text{Raven} \Rightarrow \text{Black} \quad \Leftrightarrow \quad p(\text{Black}|\text{Raven}) = 1$$

$$p(\text{Not a Raven}|\text{Not Black}) = 1 - \frac{1 - p(\text{Black}|\text{Raven})}{1 - p(\text{Black})} p(\text{Raven}) \neq p(\text{Black}|\text{Raven})$$

Most importantly: the weight of evidence (Bayes factor) is small for a red apple (Good, 1960).

Popper's Criterion of Demarcation & Bayesian Epistemology:

No need to require falsification $\Rightarrow$ the weight of evidence tells us how relevant an observation is.

Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

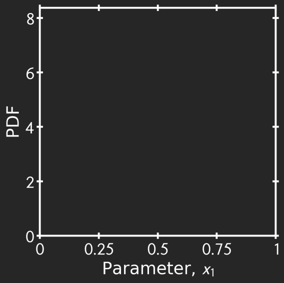
Multiple experiments increasing evidence, but single observations are meaningful.

Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

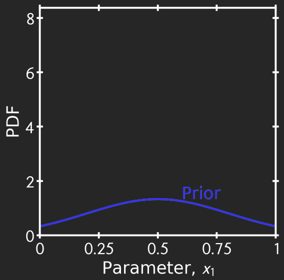
The Parsimony Principle is Hard-Coded in Bayes Factors:



Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:

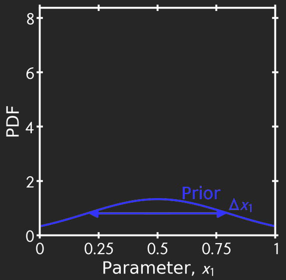


Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

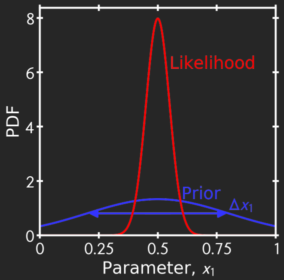
The Parsimony Principle is Hard-Coded in Bayes Factors:



Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:

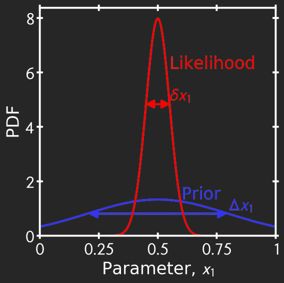


Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:

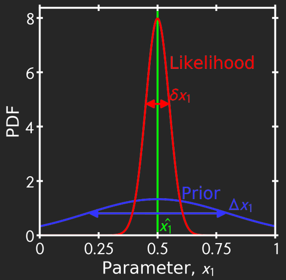


Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:

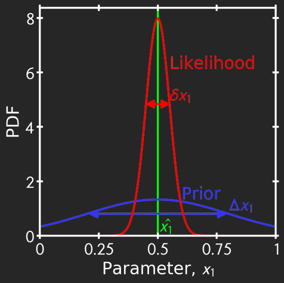


Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:



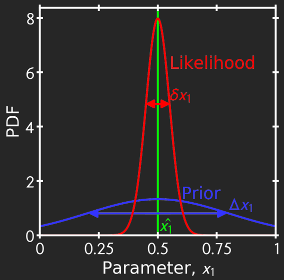
$$p(M_1|data) \quad =$$

Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:



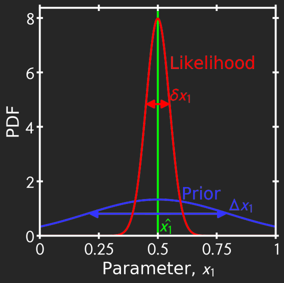
$$p(M_1|data) = \int p(data|x_1) \times p(x_1|M_1)\,\mathrm{d}x_1$$

Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:



$$p(M_1|data) = \int p(data|x_1) \times p(x_1|M_1)\,\mathrm{d}x_1$$

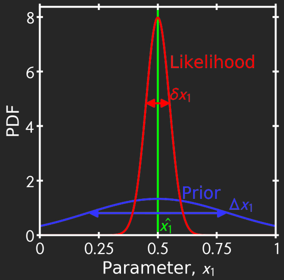
$$\propto L(\hat{x_1})\delta x_1 \times$$

Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:

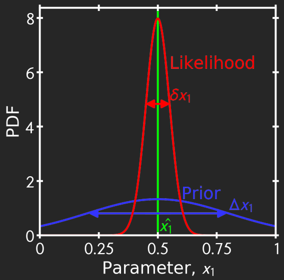


$$p(M_1|data) = \int p(data|x_1) \times p(x_1|M_1)\,\mathrm{d}x_1$$
$$\propto L(\hat{x_1})\delta x_1 \times \frac{1}{\Delta x_1}$$

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:



$$p(M_1|data) = \int p(data|x_1) \times p(x_1|M_1)\,\mathrm{d}x_1$$
$$\propto L(\hat{x_1})\delta x_1 \times \frac{1}{\Delta x_1}$$

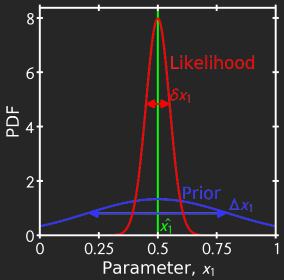
$\Rightarrow$ Bayes factor between models M_1 (1 parameter) & M_2 (2 parameters):

Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:



$$p(M_1|data) = \int p(data|x_1) \times p(x_1|M_1)\,\mathrm{d}x_1$$
$$\propto L(\hat{x_1})\delta x_1 \times \frac{1}{\Delta x_1}$$

$\Rightarrow$ Bayes factor between models M_1 (1 parameter) & M_2 (2 parameters):

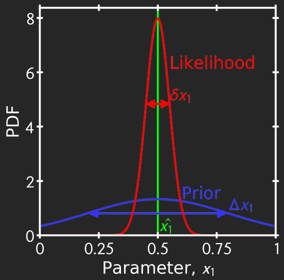
$$\frac{p(M_2|data)}{p(M_1|data)} \propto$$

Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:



$$p(M_1|data) = \int p(data|x_1) \times p(x_1|M_1)\,\mathrm{d}x_1$$

$$\propto L(\hat{x_1})\delta x_1 \times \frac{1}{\Delta x_1}$$

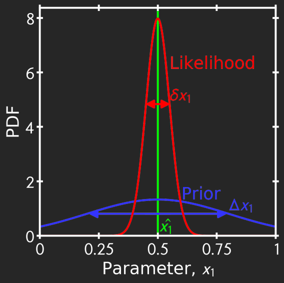
$\Rightarrow$ Bayes factor between models M_1 (1 parameter) & M_2 (2 parameters):

$$\frac{p(M_2|data)}{p(M_1|data)} \propto L(\hat{x_2}) \times \underbrace{\left(\frac{\delta x_2}{\Delta x_2}\right)}_{\ll 1}$$

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:



$$p(M_1|data) = \int p(data|x_1) \times p(x_1|M_1)\,\mathrm{d}x_1$$

$$\propto L(\hat{x_1})\delta x_1 \times \frac{1}{\Delta x_1}$$

$\Rightarrow$ Bayes factor between models M_1 (1 parameter) & M_2 (2 parameters):

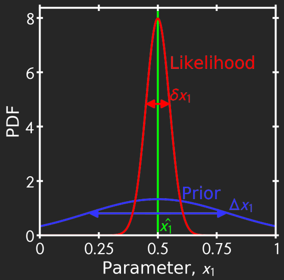
$$\frac{p(M_2|data)}{p(M_1|data)} \propto L(\hat{x_2}) \times \underbrace{\left(\frac{\delta x_2}{\Delta x_2}\right)}_{\ll 1}$$

The Prior Allows Accumulation of Knowledge:

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:



$$p(M_1|data) = \int p(data|x_1) \times p(x_1|M_1)\,\mathrm{d}x_1$$
$$\propto L(\hat{x_1})\delta x_1 \times \frac{1}{\Delta x_1}$$

$\Rightarrow$ Bayes factor between models M_1 (1 parameter) & M_2 (2 parameters):

$$\frac{p(M_2|data)}{p(M_1|data)} \propto L(\hat{x_2}) \times \underbrace{\left(\frac{\delta x_2}{\Delta x_2}\right)}_{\ll 1}$$

The Prior Allows Accumulation of Knowledge:

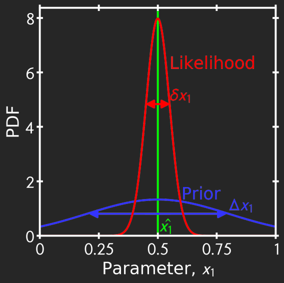
$$\underbrace{p(par)}_{\text{new prior}}$$

Reproducibility, Parsimony & Accumulation of Knowledge

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:



$$p(M_1|data) = \int p(data|x_1) \times p(x_1|M_1)\,\mathrm{d}x_1$$
$$\propto L(\hat{x_1})\delta x_1 \times \frac{1}{\Delta x_1}$$

$\Rightarrow$ Bayes factor between models M_1 (1 parameter) & M_2 (2 parameters):

$$\frac{p(M_2|data)}{p(M_1|data)} \propto L(\hat{x_2}) \times \underbrace{\left(\frac{\delta x_2}{\Delta x_2}\right)}_{\ll 1}$$

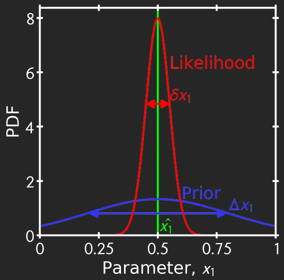
The Prior Allows Accumulation of Knowledge:

$$\underbrace{p(par)}_{\text{new prior}} \propto \underbrace{1}_{\text{initial prior}} \times$$

Reproducibility is Useful, But Not Necessary:

Multiple experiments increasing evidence, but single observations are meaningful.

The Parsimony Principle is Hard-Coded in Bayes Factors:



$$p(M_1|data) = \int p(data|x_1) \times p(x_1|M_1)\,\mathrm{d}x_1$$
$$\propto L(\hat{x_1})\delta x_1 \times \frac{1}{\Delta x_1}$$

$\Rightarrow$ Bayes factor between models M_1 (1 parameter) & M_2 (2 parameters):

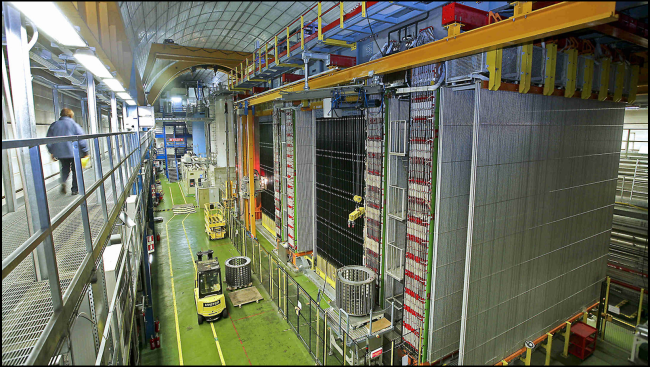
$$\frac{p(M_2|data)}{p(M_1|data)} \propto L(\hat{x_2}) \times \underbrace{\left(\frac{\delta x_2}{\Delta x_2}\right)}_{\ll 1}$$

The Prior Allows Accumulation of Knowledge:

$$\underbrace{p(par)}_{\text{new prior}} \propto \underbrace{1}_{\text{initial prior}} \times \underbrace{p(data^{(1)}|par) \times \ldots \times p(data^{(N)}|par)}_{\text{cumulated data}}$$

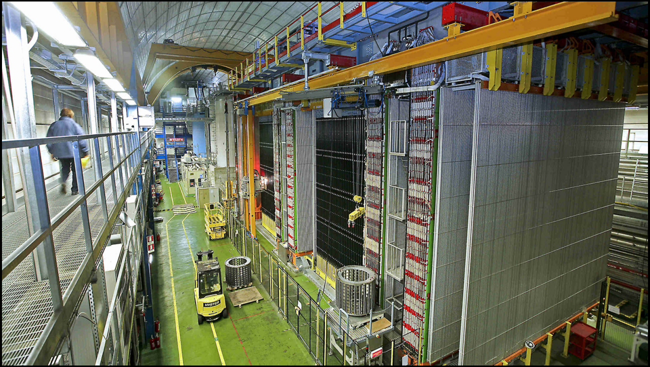
OPERA: an Experiment Refuted by Theory

OPERA: an Experiment Refuted by Theory



Oscillation Project with Emulsion-tRacking Apparatus (OPERA):

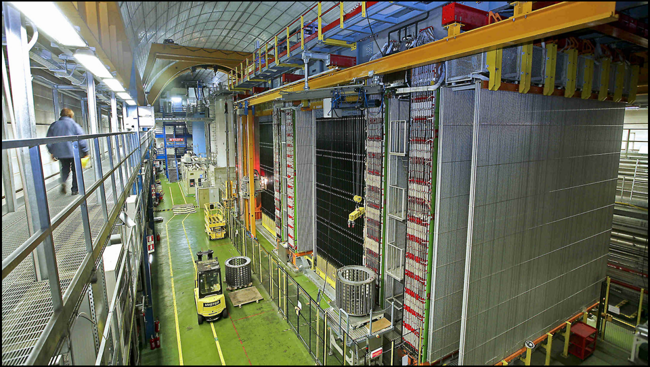
OPERA: an Experiment Refuted by Theory



Oscillation Project with Emulsion-tRacking Apparatus (OPERA):

In 2011, it detected neutrinos appearing $1 + 2 \times 10^{-5}$ times *faster than light*, at 6σ significance (frequentist analysis; OPERA coll., 2012).

OPERA: an Experiment Refuted by Theory



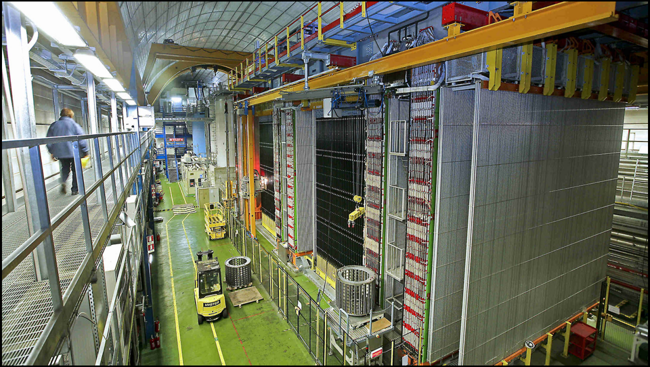
Oscillation Project with Emulsion-tRacking Apparatus (OPERA):

In 2011, it detected neutrinos appearing $1 + 2 \times 10^{-5}$ times *faster than light*, at 6σ significance (frequentist analysis; OPERA coll., 2012).

Strict Popperian/Frequentist:

$\Rightarrow$ reject special relativity.

OPERA: an Experiment Refuted by Theory



Oscillation Project with Emulsion-tRacking Apparatus (OPERA):

In 2011, it detected neutrinos appearing $1 + 2 \times 10^{-5}$ times *faster than light*, at 6σ significance (frequentist analysis; OPERA coll., 2012).

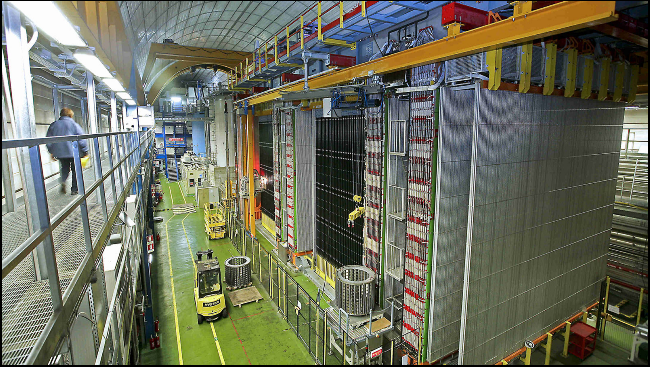
Strict Popperian/Frequentist:

$\Rightarrow$ reject special relativity.

Bayesian:

$p(v_\nu > c) \ll 1$

OPERA: an Experiment Refuted by Theory



Oscillation Project with Emulsion-tRacking Apparatus (OPERA):

In 2011, it detected neutrinos appearing $1 + 2 \times 10^{-5}$ times *faster than light*, at 6σ significance (frequentist analysis; OPERA coll., 2012).

Strict Popperian/Frequentist:

$\Rightarrow$ reject special relativity.

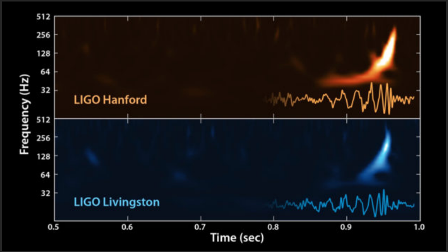
Bayesian:

$p(v_\nu > c) \ll 1 \Rightarrow$ check your cables.

Is Verifying General Relativity Useful?

Is Verifying General Relativity Useful?

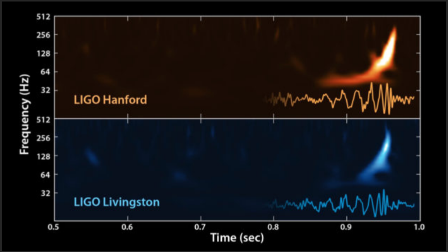
First Detection of Gravitational Waves:



(LIGO coll., 2015)

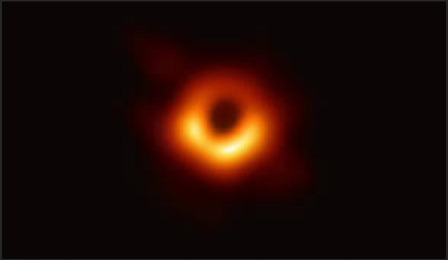
Is Verifying General Relativity Useful?

First Detection of Gravitational Waves:



(LIGO coll., 2015)

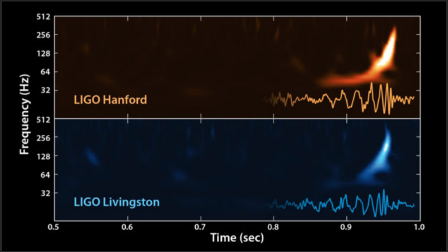
First Image of a Black Hole (M 87):



(Event Horizon coll., 2019)

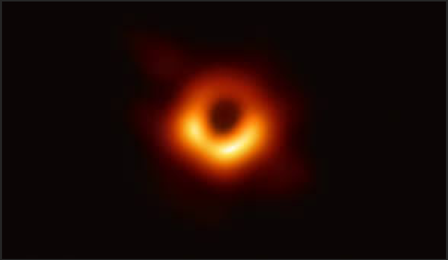
Is Verifying General Relativity Useful?

First Detection of Gravitational Waves:



(LIGO coll., 2015)

First Image of a Black Hole (M 87):

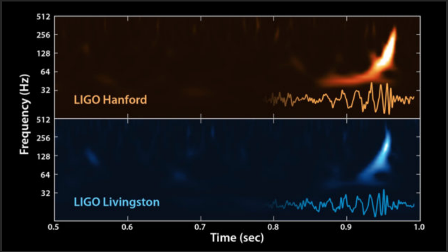


(Event Horizon coll., 2019)

Strict Popperian/Frequentist Point of View:

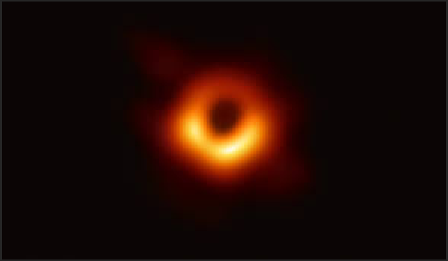
Is Verifying General Relativity Useful?

First Detection of Gravitational Waves:



(LIGO coll., 2015)

First Image of a Black Hole (M 87):



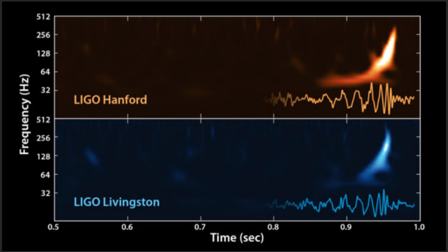
(Event Horizon coll., 2019)

Strict Popperian/Frequentist Point of View:

1. Those two experiments are *distinct*.

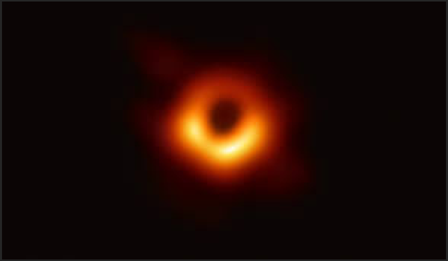
Is Verifying General Relativity Useful?

First Detection of Gravitational Waves:



(LIGO coll., 2015)

First Image of a Black Hole (M 87):



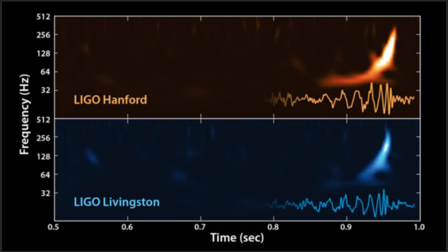
(Event Horizon coll., 2019)

Strict Popperian/Frequentist Point of View:

1. Those two experiments are *distinct*.
2. An absence of detection would not have disproven general relativity

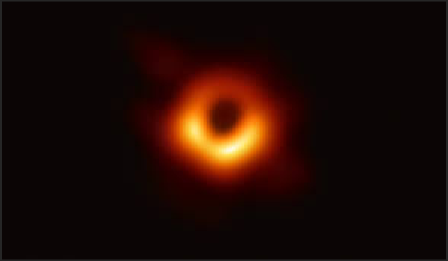
Is Verifying General Relativity Useful?

First Detection of Gravitational Waves:



(LIGO coll., 2015)

First Image of a Black Hole (M 87):



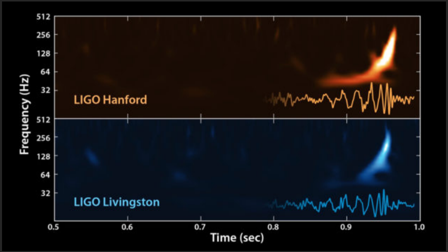
(Event Horizon coll., 2019)

Strict Popperian/Frequentist Point of View:

1. Those two experiments are *distinct*.
2. An absence of detection would not have disproven general relativity $\Rightarrow$ *no falsifiability*.

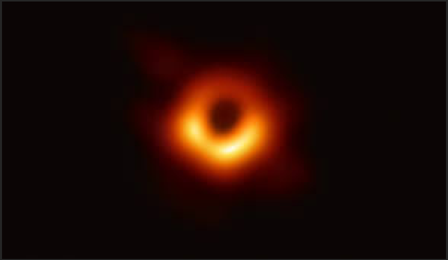
Is Verifying General Relativity Useful?

First Detection of Gravitational Waves:



(LIGO coll., 2015)

First Image of a Black Hole (M 87):



(Event Horizon coll., 2019)

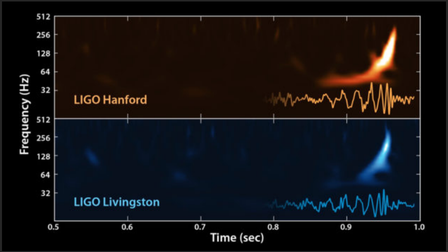
Strict Popperian/Frequentist Point of View:

1. Those two experiments are *distinct*.
2. An absence of detection would not have disproven general relativity $\Rightarrow$ *no falsifiability*.

Bayesian Point of View:

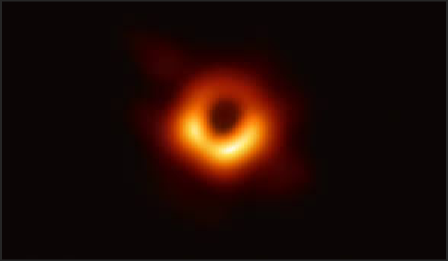
Is Verifying General Relativity Useful?

First Detection of Gravitational Waves:



(LIGO coll., 2015)

First Image of a Black Hole (M 87):



(Event Horizon coll., 2019)

Strict Popperian/Frequentist Point of View:

1. Those two experiments are *distinct*.
2. An absence of detection would not have disproven general relativity $\Rightarrow$ *no falsifiability*.

Bayesian Point of View:

Those two experiments, combined, bring a large *weight of evidence* in favor of general relativity.

Outline of the Talk

Summary & Conclusion

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:

    Bayesian: probabilities quantify our *knowledge*;

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:

   Bayesian: probabilities quantify our *knowledge*;
   Frequentist: probabilities represent the *frequency of occurence* of a repeated event.

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:
   Bayesian: probabilities quantify our *knowledge*;
   Frequentist: probabilities represent the *frequency of occurence* of a repeated event.
2. It appears there are *fundamental issues* with the frequentist approach:

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:

   Bayesian: probabilities quantify our *knowledge*;

   Frequentist: probabilities represent the *frequency of occurence* of a repeated event.

2. It appears there are *fundamental issues* with the frequentist approach:
   - p-values can lead to inconsistent results;

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:

   Bayesian: probabilities quantify our *knowledge*;

   Frequentist: probabilities represent the *frequency of occurence* of a repeated event.

2. It appears there are *fundamental issues* with the frequentist approach:
   - p-values can lead to inconsistent results;
   - the method partitions knowledge;

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:

   Bayesian: probabilities quantify our *knowledge*;
   Frequentist: probabilities represent the *frequency of occurence* of a repeated event.

2. It appears there are *fundamental issues* with the frequentist approach:
   - p-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:
   - **Bayesian:** probabilities quantify our *knowledge*;
   - **Frequentist:** probabilities represent the *frequency of occurence* of a repeated event.
2. It appears there are *fundamental issues* with the frequentist approach:
   - *p*-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.
3. After a century of frequentist supremacy, Bayesianism emerged victorious.

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:
   Bayesian: probabilities quantify our *knowledge*;
   Frequentist: probabilities represent the *frequency of occurence* of a repeated event.
2. It appears there are *fundamental issues* with the frequentist approach:
   - p-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.
3. After a century of frequentist supremacy, Bayesianism emerged victorious.

Epistemological Insights from Bayesianism:

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:
   - **Bayesian:** probabilities quantify our *knowledge*;
   - **Frequentist:** probabilities represent the *frequency of occurence* of a repeated event.
2. It appears there are *fundamental issues* with the frequentist approach:
   - p-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.
3. After a century of frequentist supremacy, Bayesianism emerged victorious.

Epistemological Insights from Bayesianism:

Popper's influential scientific method relies on *falsifiability & reproducibility*. It has a frequentist frame of mind.

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:
   - **Bayesian:** probabilities quantify our *knowledge*;
   - **Frequentist:** probabilities represent the *frequency of occurence* of a repeated event.
2. It appears there are *fundamental issues* with the frequentist approach:
   - *p*-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.
3. After a century of frequentist supremacy, Bayesianism emerged victorious.

Epistemological Insights from Bayesianism:

Popper's influential scientific method relies on *falsifiability & reproducibility*. It has a frequentist frame of mind.

Bayesian alternative addresses the problems of Popper's epistemology:

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:
   - **Bayesian:** probabilities quantify our *knowledge*;
   - **Frequentist:** probabilities represent the *frequency of occurence* of a repeated event.
2. It appears there are *fundamental issues* with the frequentist approach:
   - p-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.
3. After a century of frequentist supremacy, Bayesianism emerged victorious.

Epistemological Insights from Bayesianism:

Popper's influential scientific method relies on *falsifiability & reproducibility*. It has a frequentist frame of mind.

Bayesian alternative addresses the problems of Popper's epistemology:

- No need for falsifiability $\Rightarrow$ *weight of evidence* quantifies data relevance;

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:

   Bayesian: probabilities quantify our *knowledge*;

   Frequentist: probabilities represent the *frequency of occurence* of a repeated event.

2. It appears there are *fundamental issues* with the frequentist approach:
   - p-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.

3. After a century of frequentist supremacy, Bayesianism emerged victorious.

Epistemological Insights from Bayesianism:

Popper's influential scientific method relies on *falsifiability & reproducibility*. It has a frequentist frame of mind.

Bayesian alternative addresses the problems of Popper's epistemology:

- No need for falsifiability $\Rightarrow$ *weight of evidence* quantifies data relevance;
- No need for reproducibility $\Rightarrow$ can account for *sparser, unique constraints*;

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:

   Bayesian: probabilities quantify our *knowledge*;

   Frequentist: probabilities represent the *frequency of occurence* of a repeated event.

2. It appears there are *fundamental issues* with the frequentist approach:
   - *p*-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.

3. After a century of frequentist supremacy, Bayesianism emerged victorious.

Epistemological Insights from Bayesianism:

Popper's influential scientific method relies on *falsifiability & reproducibility*. It has a frequentist frame of mind.

Bayesian alternative addresses the problems of Popper's epistemology:

- No need for falsifiability $\Rightarrow$ *weight of evidence* quantifies data relevance;
- No need for reproducibility $\Rightarrow$ can account for *sparser, unique constraints*;
- Parsimony $\Rightarrow$ Bayes factors;

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:
   - **Bayesian:** probabilities quantify our *knowledge*;
   - **Frequentist:** probabilities represent the *frequency of occurence* of a repeated event.
2. It appears there are *fundamental issues* with the frequentist approach:
   - *p*-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.
3. After a century of frequentist supremacy, Bayesianism emerged victorious.

Epistemological Insights from Bayesianism:

Popper's influential scientific method relies on *falsifiability & reproducibility*. It has a frequentist frame of mind.

Bayesian alternative addresses the problems of Popper's epistemology:

- No need for falsifiability $\Rightarrow$ *weight of evidence* quantifies data relevance;
- No need for reproducibility $\Rightarrow$ can account for *sparser, unique constraints*;
- Parsimony $\Rightarrow$ Bayes factors;
- Allows *accumulation of knowledge* $\Rightarrow$ informative prior;

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:
   - **Bayesian:** probabilities quantify our *knowledge*;
   - **Frequentist:** probabilities represent the *frequency of occurence* of a repeated event.
2. It appears there are *fundamental issues* with the frequentist approach:
   - p-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.
3. After a century of frequentist supremacy, Bayesianism emerged victorious.

Epistemological Insights from Bayesianism:

Popper's influential scientific method relies on *falsifiability & reproducibility*. It has a frequentist frame of mind.

Bayesian alternative addresses the problems of Popper's epistemology:

- No need for falsifiability $\Rightarrow$ *weight of evidence* quantifies data relevance;
- No need for reproducibility $\Rightarrow$ can account for *sparser, unique constraints*;
- Parsimony $\Rightarrow$ Bayes factors;
- Allows *accumulation of knowledge* $\Rightarrow$ informative prior;
- Allows probabilistic *induction*.

Summary & Conclusion

The Opposition Between Bayesianism & Frequentism:

1. There are two competing epistemological conceptions of probability:
   - **Bayesian:** probabilities quantify our *knowledge*;
   - **Frequentist:** probabilities represent the *frequency of occurence* of a repeated event.
2. It appears there are *fundamental issues* with the frequentist approach:
   - p-values can lead to inconsistent results;
   - the method partitions knowledge;
   - it is less flexible than Bayesianism.
3. After a century of frequentist supremacy, Bayesianism emerged victorious.

Epistemological Insights from Bayesianism:

Popper's influential scientific method relies on *falsifiability & reproducibility*. It has a frequentist frame of mind.

Bayesian alternative addresses the problems of Popper's epistemology:

- No need for falsifiability $\Rightarrow$ *weight of evidence* quantifies data relevance;
- No need for reproducibility $\Rightarrow$ can account for *sparser, unique constraints*;
- Parsimony $\Rightarrow$ Bayes factors;
- Allows *accumulation of knowledge* $\Rightarrow$ informative prior;
- Allows probabilistic *induction*.

This is already the way we think (at least qualitatively), because it is the way *our brain works*.