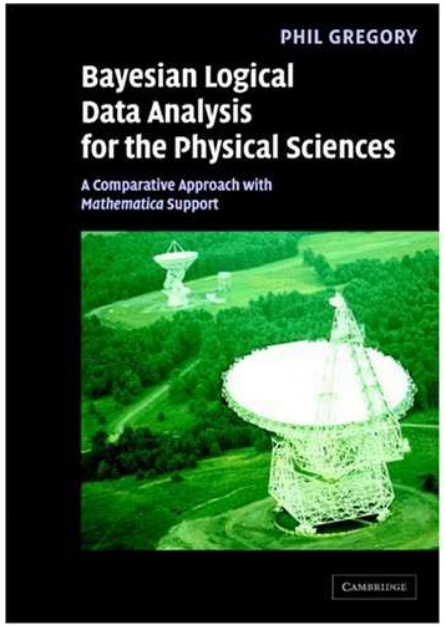
1. Introduction and Theoretical Foundations

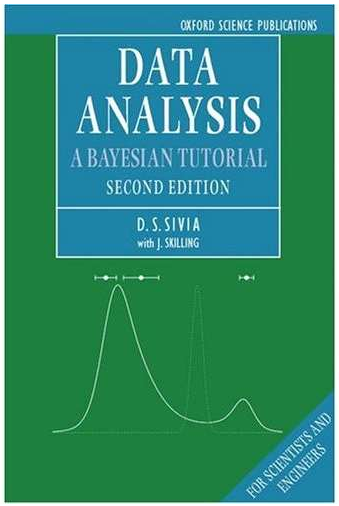
Reasonable thinking?…



PREFACE

The goal of science is to unlock nature's secrets…Our understanding comes through the development of theoretical models capable of explaining the existing observations as well as making testable predictions…Statistical inference provides a means for assessing the plausibility of one or more competing models, and estimating the model parameters and their uncertainities. These topics are commonly referred to as "data analysis".

The most we can hope to do is to make the best inference based on the experimental data and any prior knowledge that we have available.



Reasonable thinking?...

Herodotus, c.500 BC

"A decision was wise, even though it led to disastrous consequences, if the evidence at hand indicated it was the best one to make; and a decision was foolish, even though it led to the happiest possible consequences, if it was unreasonable to expect those consequences"

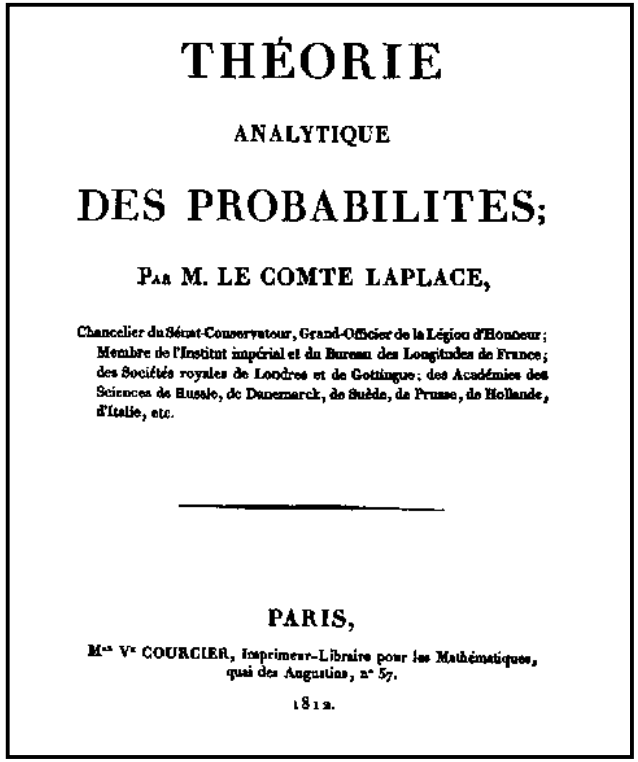
Reasonable thinking?...

Herodotus, c.500 BC

"A decision was wise, even though it led to disastrous consequences, if the evidence at hand indicated it was the best one to make; and a decision was foolish, even though it led to the happiest possible consequences, if it was unreasonable to expect those consequences"

Pierre-Simon Laplace
(1749 – 1827)

"Probability theory is nothing but common sense reduced to calculation"



THÉORIE

ANALYTIQUE

DES PROBABILITES;

PAR M. LE COMTE LAPLACE,

Chancelier du Sénat-Conservateur, Grand-Officier de la Légion d'Honneur; Membre de l'Institut impérial et du Bureau des Longitudes de France; des Sociétés royales de Londres et de Gottingue; des Académies des Sciences de Russie, de Danemarck, de Suède, de Prusse, de Hollande, d'Italie, etc.

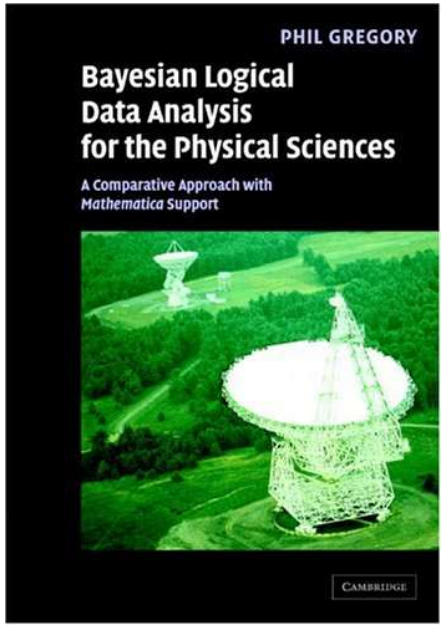
PARIS,

Mme Ve COURCIER, Imprimeur-Libraire pour les Mathématiques, quai des Augustins, n° 57.

1812.

1. Introduction and Theoretical Foundations

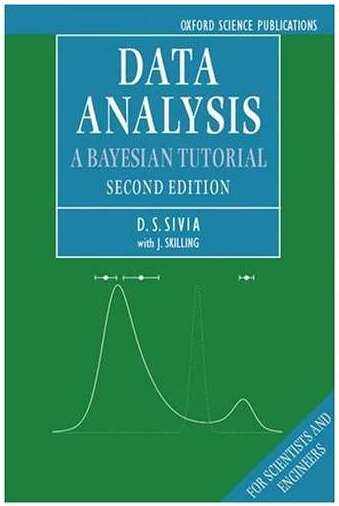
Plausible reasoning?…



PREFACE

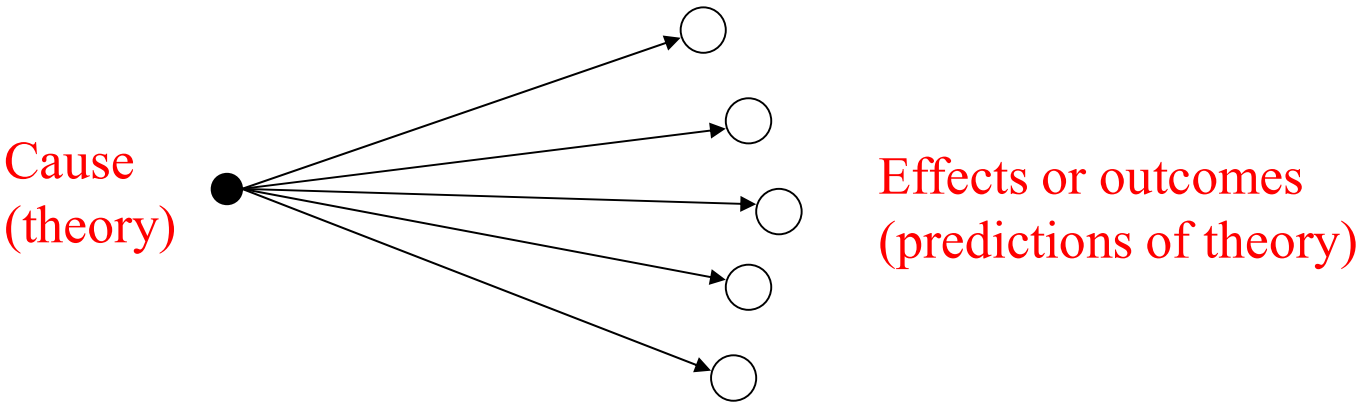
The goal of science is to unlock nature's secrets…Our understanding comes through the development of theoretical models capable of explaining the existing observations as well as making testable predictions…Statistical inference provides a means for assessing the plausibility of one or more competing models, and estimating the model parameters and their uncertainities. These topics are commonly referred to as "data analysis".

The most we can hope to do is to make the best inference based on the experimental data and any prior knowledge that we have available.



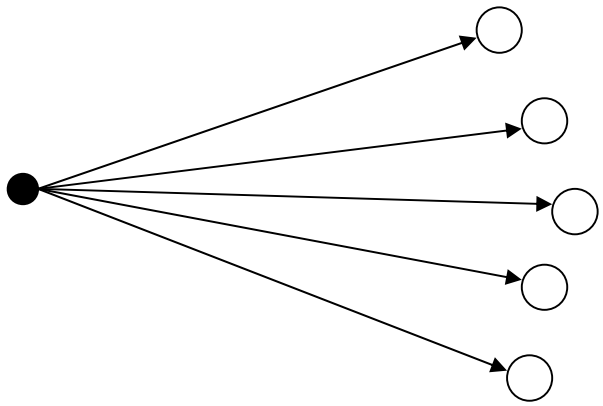
We need to think about the difference between **deductive** and **inductive** logic

Deductive logic



Deductive logic

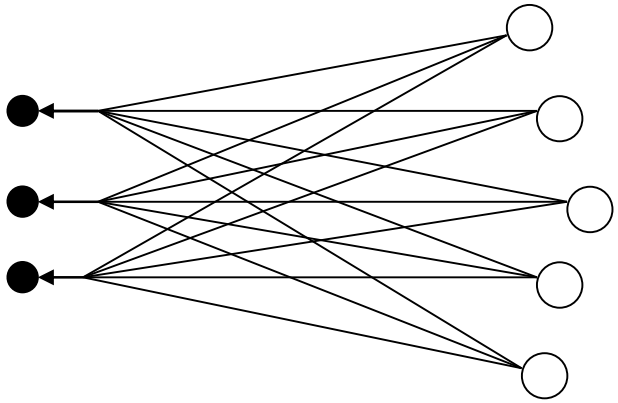
Cause
(theory)



Effects or outcomes
(predictions of theory)

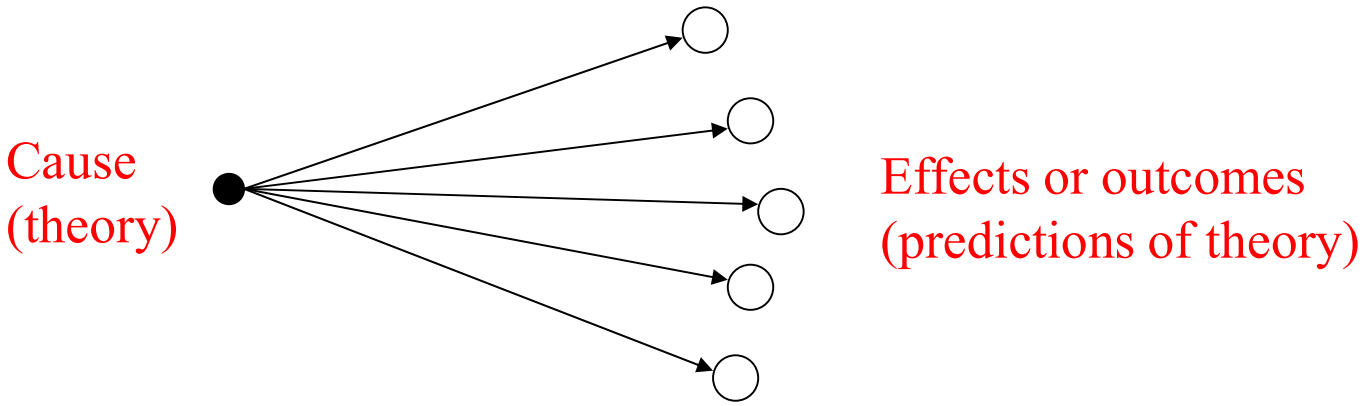
Inductive logic

Possible causes
(competing theories
or models)

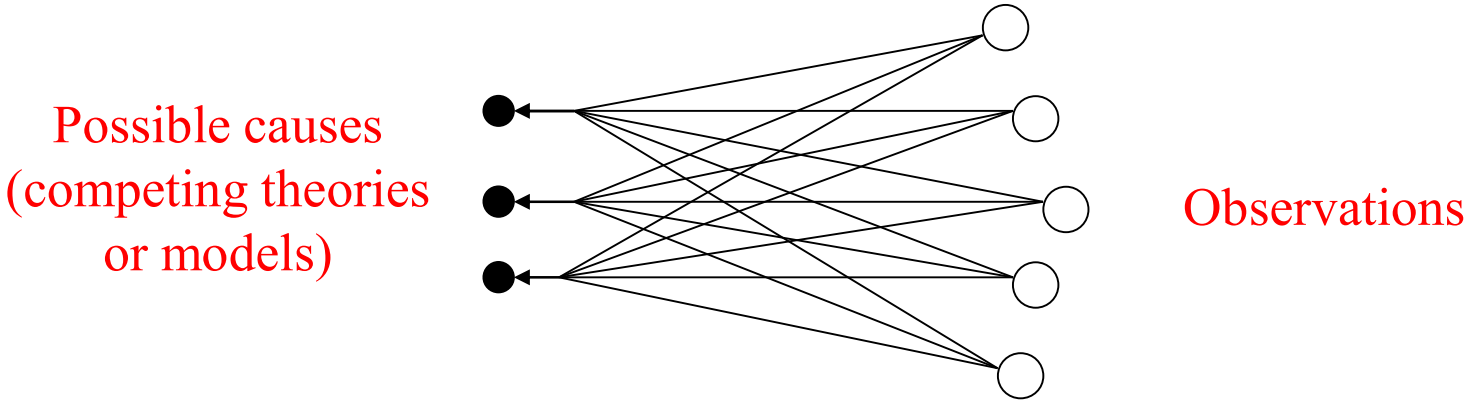


Observations

Deductive logic

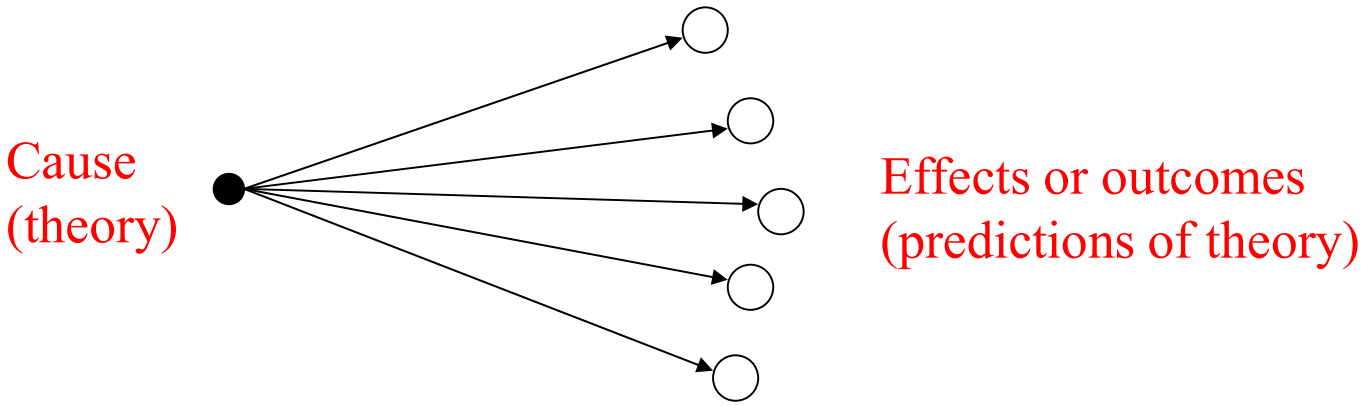


Inductive logic

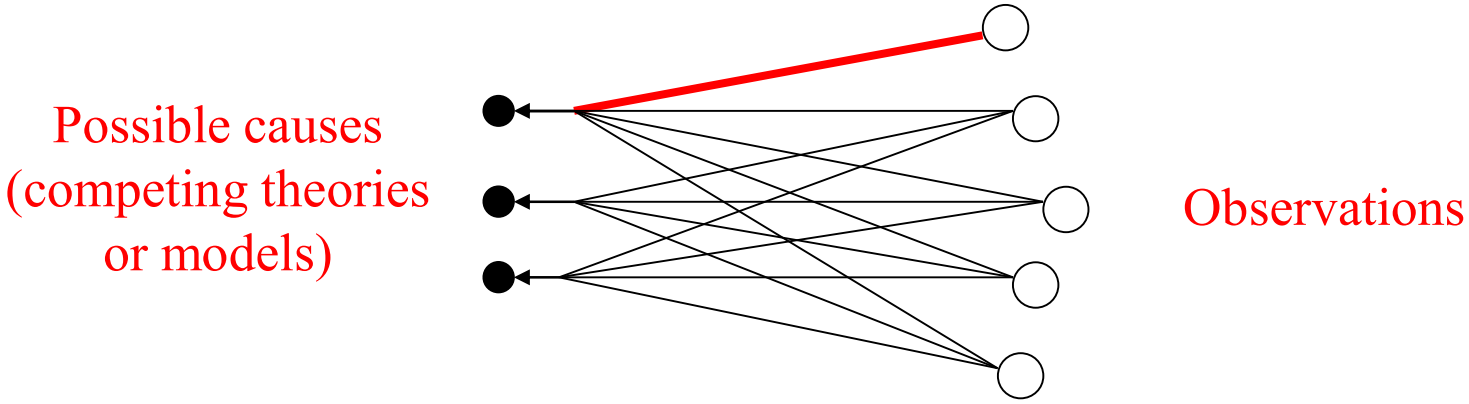


How do we decide which model is most **plausible**?

Deductive logic



Inductive logic



How do we decide which model is most **plausible**?

An example of deductive logic

Statement A:	All red-haired students drink Irn Bru
Statement B:	Student X has red hair
Statement C:	Student X drinks Irn Bru

An example of deductive logic

Statement A:	All red-haired students drink Irn Bru
Statement B:	Student X has red hair
Statement C:	Student X drinks Irn Bru

Let's suppose that A is true. (Our theory).

- If B is true, then C is true
- If C is false, then B is false

An example of deductive logic

Statement A:	All red-haired students drink Irn Bru
Statement B:	Student X has red hair
Statement C:	Student X drinks Irn Bru

Let's suppose that A is true. (Our theory).

- If B is true, then C is true
- If C is false, then B is false

C is a logical consequence of A and B

If we set 'true' = 1 and 'false' = 0, we can use the rules of George Boole (1854) to carry out logical operations.

We define

Negation:	$\overline{A}$	'A is false'
Logical product:	AB	'both A and B are true'
Logical sum:	$A+B$	'at least one of A or B is true'

Then

$$A(B+C) = AB + AC$$

$$A + AB = A$$

$$A + \overline{A} = 1$$

$$A + BC = (A+B)(A+C)$$

$$A\overline{A} = 0 \qquad \text{etc}$$

An example of inductive logic

Statement A:	All red-haired students drink Irn Bru
Statement B:	Student X has red hair
Statement C:	Student X drinks Irn Bru

What can we say about B if A and C are true?...

(Statement A didn't say that all students who drink Irn Bru have red hair)

An example of inductive logic

Statement A:	All red-haired students drink Irn Bru
Statement B:	Student X has red hair
Statement C:	Student X drinks Irn Bru

What can we say about B if A and C are true?...

(Statement A didn't say that all students who drink Irn Bru have red hair)

We might say, however

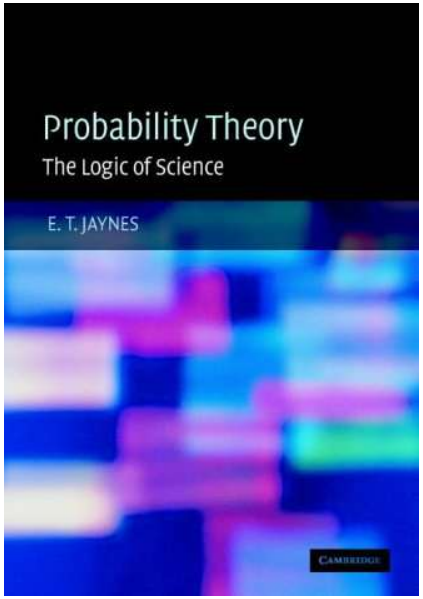
- If C is true, then B is more plausible

In the 1940s and 50s Cox, Polya and Jaynes formalised the mathematics of inductive logic as plausible reasoning

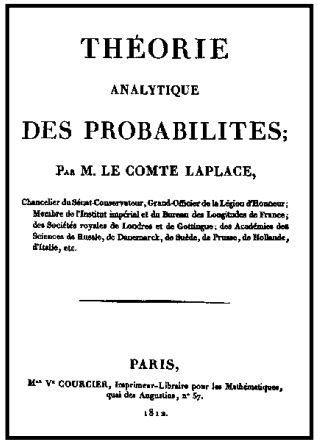
- If we assign degrees of plausibility a real number between 0 and 1, then the rules for combining and operating on inductive logical statements are identical to those for deductive logic $\longrightarrow$ Boolean algebra.

In the 1940s and 50s Cox, Polya and Jaynes formalised the mathematics of inductive logic as plausible reasoning

- If we assign degrees of plausibility a real number between 0 and 1, then the rules for combining and operating on inductive logical statements are identical to those for deductive logic ⟶ Boolean algebra.



Ed Jaynes
(1922 – 1998)



THÉORIE

ANALYTIQUE

DES PROBABILITES;

PAR M. LE COMTE LAPLACE,

Chancelier du Sénat-Conservateur, Grand-Officier de la Légion d'Honneur; Membre de l'Institut impérial et du Bureau des Longitudes de France; des Sociétés royales de Londres et de Gottingue; des Académies des Sciences de Russie, de Danemarck, de Suède, de Prusse, de Hollande, d'Italie, etc.

PARIS,

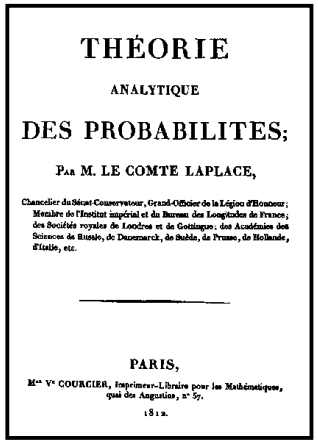
Mme Ve COURCIER, Imprimeur-Libraire pour les Mathématiques, quai des Augustins, n° 57.

1812.

Laplace (1812)

Mathematical framework for probability as a basis for plausible reasoning:

Probability measures our degree of belief that something is true



THÉORIE

ANALYTIQUE

DES PROBABILITES;

PAR M. LE COMTE LAPLACE,

Chancelier du Sénat-Conservateur, Grand-Officier de la Légion d'Honneur; Membre de l'Institut impérial et du Bureau des Longitudes de France; des Sociétés royales de Londres et de Gottingue; des Académies des Sciences de Russie, de Danemarck, de Suède, de Prusse, de Hollande, d'Italie, etc.

PARIS,

Mme Ve COURCIER, Imprimeur-Libraire pour les Mathématiques, quai des Augustins, n° 57.

1812.

Laplace (1812)

Mathematical framework for probability as a basis for plausible reasoning:

Probability measures our degree of belief that something is true

$$\text{Prob}(X) = 1 \quad \Rightarrow \quad \text{we are } \textit{certain} \text{ that } X \text{ is true}$$

$$\text{Prob}(X) = 0 \quad \Rightarrow \quad \text{we are } \textit{certain} \text{ that } X \text{ is false}$$

Our degree of belief always depends on the available background information:

We write $\boxed{\text{Prob}(X \mid I)}$ "Probability that X is true, given I"

Background information

Vertical line denotes conditional probability:

our state of knowledge about X is *conditioned* by background info, I

Rules for combining probabilities

$$\boxed{p(X \mid I) \quad + \quad p(\overline{X} \mid I) \quad = \quad 1}$$

$\overline{X}$ denotes the proposition that X is false

Note: the background information is the *same* in both cases

Rules for combining probabilities

$$\boxed{p(X,Y\,|\,I) \quad = \quad p(X\,|\,Y,I)\times p(Y\,|\,I)}$$

X,Y denotes the proposition that X **and** Y are true

Rules for combining probabilities

$$\boxed{p(X,Y\,|\,I) \quad = \quad p(X\,|\,Y,I)\times p(Y\,|\,I)}$$

X,Y denotes the proposition that X **and** Y are true

$p(X\,|\,Y,I)$ = Prob(X is true, given Y is true)

$p(Y\,|\,I)$ = Prob(Y is true, irrespective of X)

Also

$$\boxed{p(Y,X\,|\,I) \quad = \quad p(Y\,|\,X,I)\times p(X\,|\,I)}$$

Note

$$p(X\,|\,Y,I_1) \quad \neq \quad p(X\,|\,Y,I_2)$$

Also

$$\boxed{p(X\,|\,Y,I) \quad \neq \quad p(Y\,|\,X,I)}$$

theguardian

Guantánamo Bay files: Casio wristwatch 'the sign of al-Qaida'

Casio F-91W, a cheap digital watch sold around the world, was taken as evidence of detainees having bomb-making training

• Read the military briefing on the Casio F-91W watch

Follow James Ball by email BETA

James Ball
The Guardian, Monday 25 April 2011

The Casio F-91W wristwatch, regarded by Guantánamo Bay interrogators as a sign of al-Qaida involvement.

It is cheap, basic and widely available around the world. Yet the Casio F-91W digital watch was declared to be "the sign of al-Qaida" and a contributing factor to continued detention of prisoners by the analysts stationed at Guantánamo Bay.

Briefing documents used to train staff in assessing the threat level of new detainees advise that possession of the F-91W – available online for as little as £4 – suggests the wearer has been trained in bomb making by al-Qaida in Afghanistan.

Article history

World news
The Guantánamo files · Guantánamo Bay · Al-Qaida

More news

More on this story

Guantánamo files lift lid on world's most controversial prison
• Innocent people interrogated for years on slimmest pretexts
• Children, elderly and mentally ill among those wrongfully held
• 172 prisoners remain, some with no prospect of trial or release

Al-Qaida assassin 'worked for MI6'

Files rewrite story of Bin Laden's Tora Bora escape

From

$$\boxed{p(Y,X\,|\,I) \quad = \quad p(Y\,|\,X,I)\times p(X\,|\,I)}$$

and

$$p(Y,X\,|\,I) \quad = \quad p(X,Y\,|\,I)$$

We have

$$\boxed{p(Y\,|\,X,I) \quad = \quad \frac{p(X\,|\,Y,I)\times p(Y\,|\,I)}{p(X\,|\,I)}}$$

Bayes' theorem:

$$p(Y \mid X, I) = \frac{p(X \mid Y, I) \times p(Y \mid I)}{p(X \mid I)}$$

Laplace rediscovered work of Rev. Thomas Bayes (1763)

Bayesian Inference

Thomas Bayes
(1702 – 1761 AD)

Bayes' theorem:

$$p(Y \mid X, I) = \frac{p(X \mid Y, I) \times p(Y \mid I)}{p(X \mid I)}$$

$$p(\text{model} \mid \text{data}, I) = \frac{p(\text{data} \mid \text{model}, I) \times p(\text{model} \mid I)}{p(\text{data} \mid I)}$$

Bayes' theorem:

$$p(Y \mid X, I) = \frac{p(X \mid Y, I) \times p(Y \mid I)}{p(X \mid I)}$$

Posterior

Likelihood

Prior

$$p(\text{model} \mid \text{data}, I) = \frac{p(\text{data} \mid \text{model}, I) \times p(\text{model} \mid I)}{p(\text{data} \mid I)}$$

Evidence

Bayes' theorem:

$$p(Y \mid X, I) = \frac{p(X \mid Y, I) \times p(Y \mid I)}{p(X \mid I)}$$

Posterior

Likelihood

Prior

$$p(\text{model} \mid \text{data}, I) = \frac{p(\text{data} \mid \text{model}, I) \times p(\text{model} \mid I)}{p(\text{data} \mid I)}$$

Evidence

We can calculate these terms

Bayes' theorem:

$$p(Y \mid X, I) = \frac{p(X \mid Y, I) \times p(Y \mid I)}{p(X \mid I)}$$

Posterior

Likelihood

Prior

$$p(\text{model} \mid \text{data}, I) \propto p(\text{data} \mid \text{model}, I) \times p(\text{model} \mid I)$$

What we know now

Influence of our observations

What we knew before

Bayesian probability theory is simultaneously a very old and a very young field:

Old : original interpretation of Bernoulli, Bayes, Laplace…

Young: 'state of the art' in data analysis

But BPT was rejected for several centuries.

Bayesian probability theory is simultaneously a very old and a very young field:

Old : original interpretation of Bernoulli, Bayes, Laplace…

Young: 'state of the art' in data analysis

But BPT was rejected for several centuries.

Probability $\equiv$ degree of belief was seen as too subjective

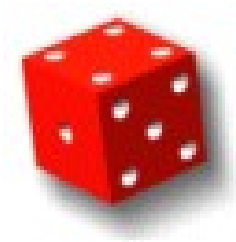
$\Downarrow$

Frequentist approach

Probability = 'long run relative frequency' of an event

in principle, it was thought, can be measured objectively

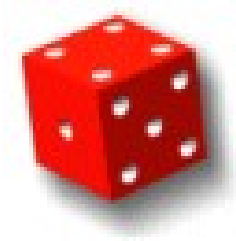
e.g. rolling a die.



What is $p(1)$?

Probability = 'long run relative frequency' of an event

in principle, it was thought, can be measured objectively

e.g. rolling a die.  What is $p(1)$?

If die is 'fair' we expect $p(1) = p(2) = \ldots = p(6) = \frac{1}{6}$

These probabilities are fixed (but unknown) numbers.

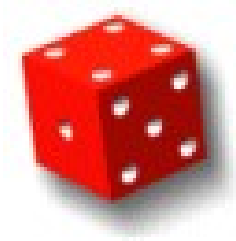
Can imagine rolling die M times.

Number rolled is a random variable – different outcome each time.

Probability = 'long run relative frequency' of an event

in principle, it was thought, can be measured objectively

e.g. rolling a die.



What is $p(1)$?

If die is 'fair' we expect $p(1) = p(2) = \ldots = p(6) = \dfrac{1}{6}$

These probabilities are fixed (but unknown) numbers.

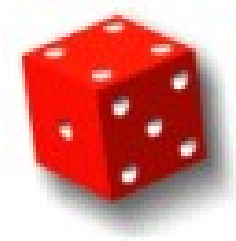
We define $p(1) = \lim_{M \to \infty} \dfrac{n(1)}{M}$

If $p(1) = \frac{1}{6}$ die is 'fair'

Probability = 'long run relative frequency' of an event

in principle, it was thought, can be measured objectively

e.g. rolling a die.



What is $p(1)$?

If die is 'fair' we expect $p(1) = p(2) = \ldots = p(6) = \frac{1}{6}$

These probabilities are fixed (but unknown) numbers.

But objectivity is an illusion:

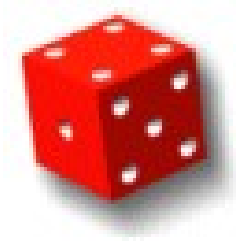
$$p(1) = \lim_{M \to \infty} \frac{n(1)}{M}$$

assumes each outcome equally likely

(i.e. equally probable)

Probability = 'long run relative frequency' of an event

in principle, it was thought, can be measured objectively

e.g. rolling a die.  What is $p(1)$?

If die is 'fair' we expect $p(1) = p(2) = \ldots = p(6) = \frac{1}{6}$

These probabilities are fixed (but unknown) numbers.

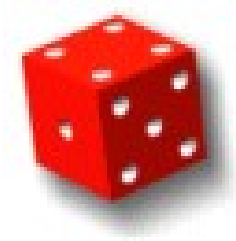
But objectivity is an illusion:

Also assumes infinite series of *identical* trials;

why can't probabilities change?

Probability = 'long run relative frequency' of an event

in principle, it was thought, can be measured objectively

e.g. rolling a die.  What is $p(1)$?

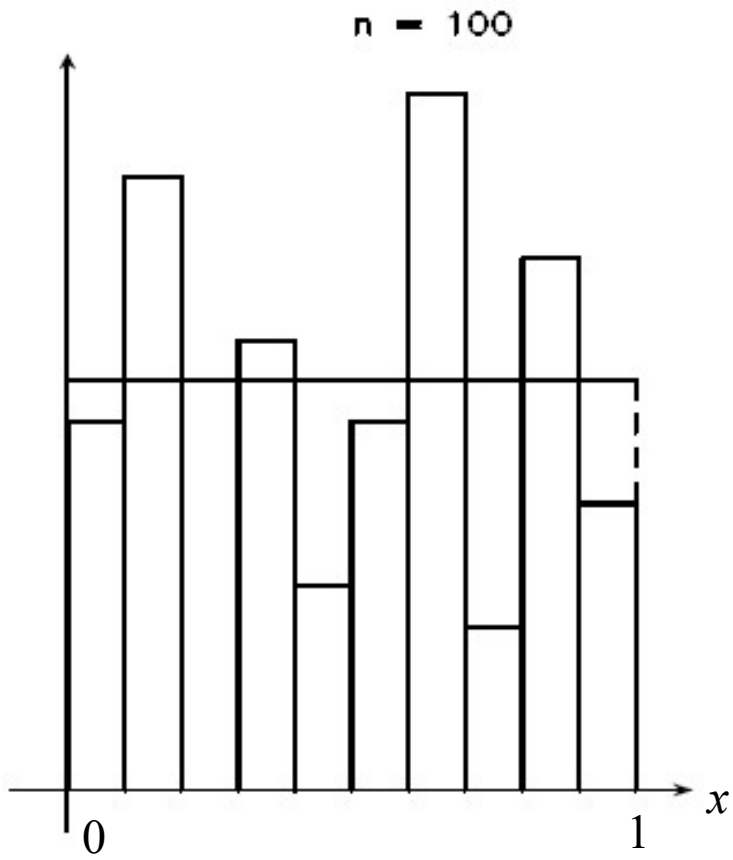
If die is 'fair' we expect $p(1) = p(2) = \ldots = p(6) = \frac{1}{6}$

These probabilities are fixed (but unknown) numbers.

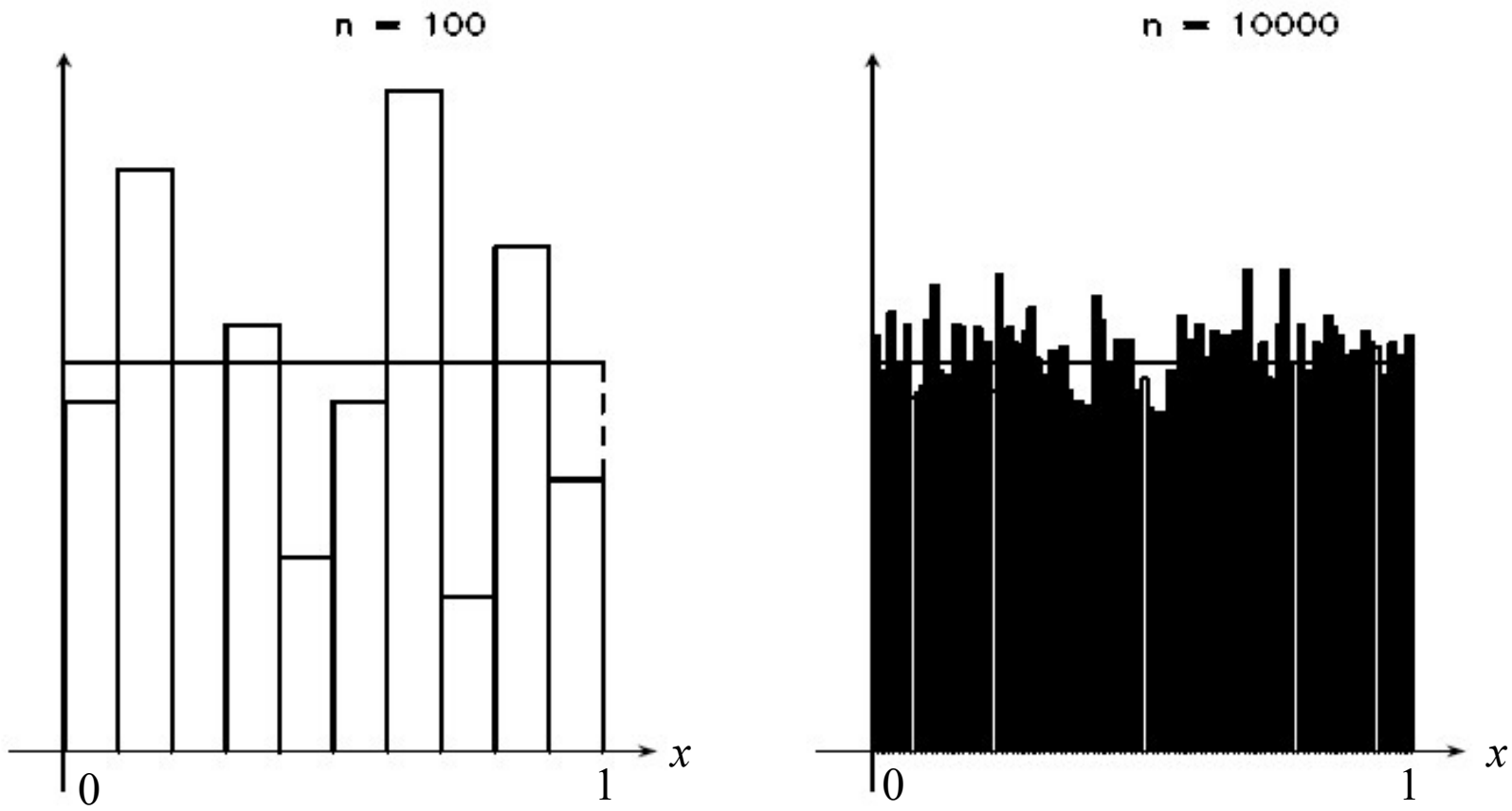
But objectivity is an illusion:

What can we say about the fairness of the die after (say) 5 rolls, or 10, or 100 ?

In the frequentist approach, a lot of mathematical machinery is defined to let us address this type of question. (See later)



In the frequentist approach, a lot of mathematical machinery is defined to let us address this type of question. (See later)

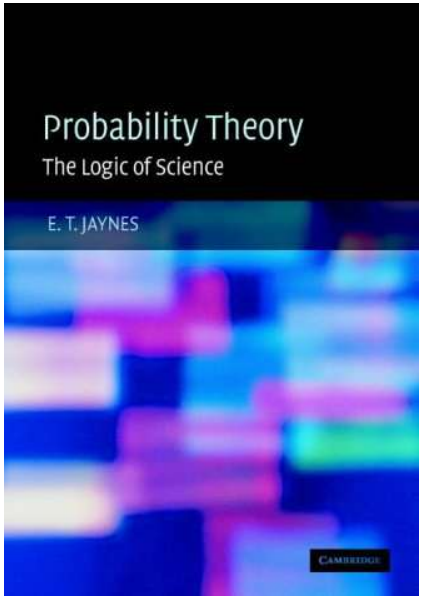


Bayesian versus Frequentist statistics: Who is right?

Frequentists are correct to worry about subjectiveness of assigning probabilities – Bayesians worry about this too!

Bayesian versus Frequentist statistics: Who is right?

Frequentists are correct to worry about subjectiveness of assigning probabilities – Bayesians worry about this too!

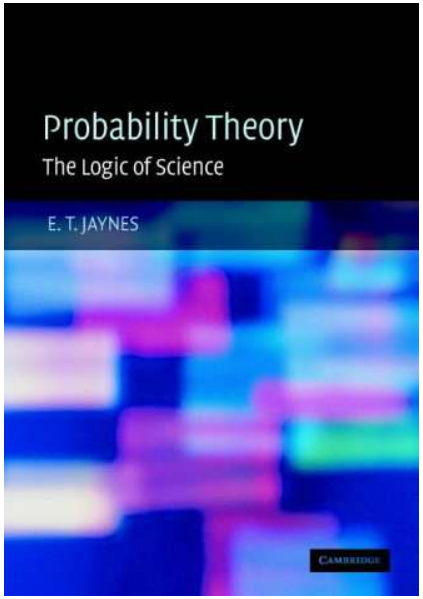


Ed Jaynes
(1922 – 1998)

Probability *is* subjective;
it depends on the available information

Bayesian versus Frequentist statistics: Who is right?

Frequentists are correct to worry about subjectiveness of assigning probabilities – Bayesians worry about this too!



Ed Jaynes
(1922 – 1998)

Probability *is* subjective;
it depends on the available information

Subjective ≠ arbitrary

Given the *same* background information, two observers should assign the *same* probabilities