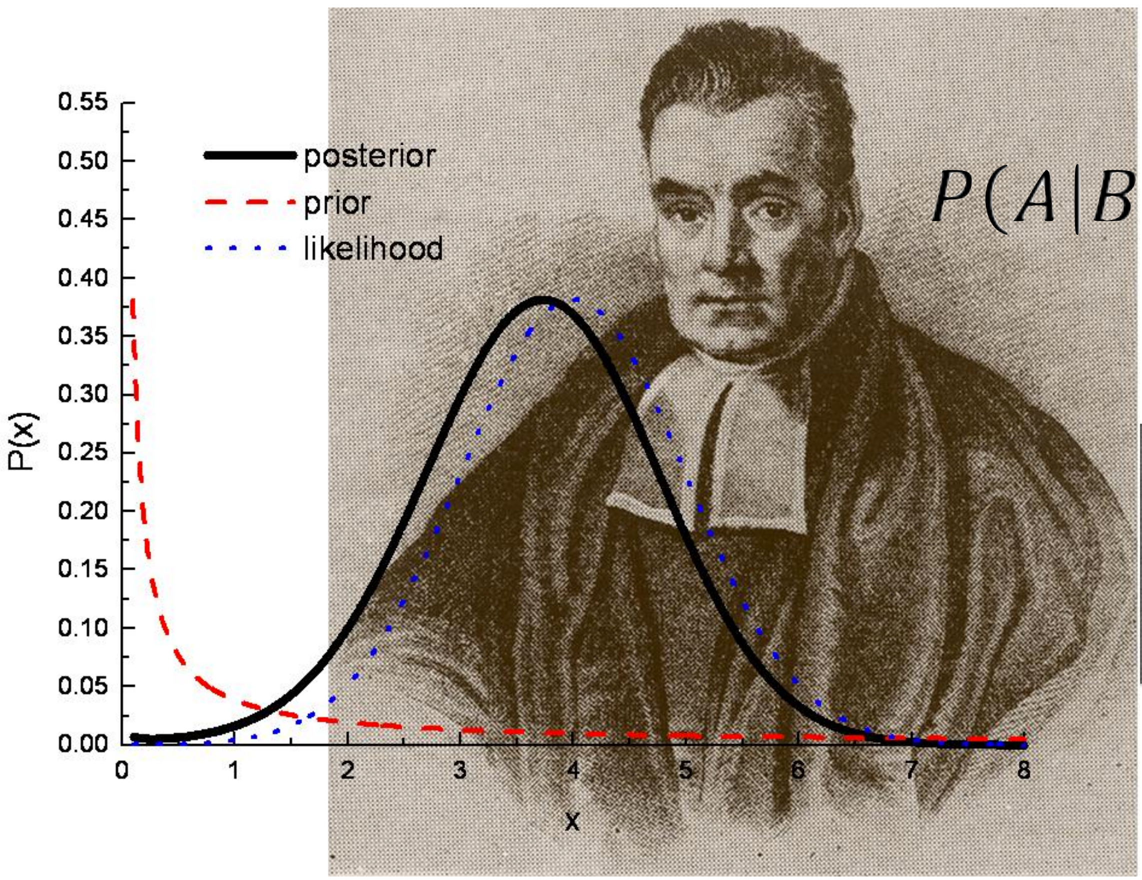
NOVEMBER 23RD 2020

BAYESIAN DATA ANALYSIS

Alexandre Herbay



$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Based on the data, what should we believe in?

Based on the data, what should we believe in?

Should I believe that the tossed coin is fair if it comes up heads in 7 of 10 flips?

Should I believe that I have COVID-19 when the test comes back positive?

Should I believe that she loves me when the daisy has 17 petals?

Our beliefs can be modified when we have data

Bayesian data analysis are techniques for making inferences from data to uncertain beliefs.

Applications in

evolution
machine learning
epidemiology
psychology
brain imaging
visual perception
Sherlock Holmes
forensic science
genetics
image processing
linguistics
human object recognition
cosmology
ecology

OUTLINE

1. Different approaches to probability
2. Bayes' rule and bayesian inference
3. Bayesian vs. frequentist inference

1. Different approaches to probability

PROBABILITY

PROBABILITY

What is a probability ?

Probability can be seen as the **long-run relative frequency** of each possible event
= potential frequency of an outcome

Probability can be seen as the **measure of uncertainty**, or **degree of belief** in the occurrence of an event
= plausibility of an outcome.

PROBABILITY

What is a probability ?

“[...] whether the probabilities should only refer to data and be based on frequency

or ***whether they should also apply to hypotheses and be regarded as measures of beliefs.****”*

(Lindley, 1993)

CONCRETE EXAMPLES

Flipping a coin : probability of a « head » result

Frequentist Probability

Make a large number of « identical » flips

Count the relative frequency of heads

Bayesian Probability

Uncertainty on the head outcome given the uncertainty on the initial conditions of the launch

CONCRETE EXAMPLES

Probability of snow today

Frequentist Probability

If we can imagine a population of many « today », repeat today many times, count the frequency of snowy days.

Bayesian Probability

Uncertainty : how certain we are that it will snow today given the initial meteorological conditions « 30% chances »

Tweet

Will Gervais @wgervais · Nov 20

TWITTER HIVEMIND.

My 9 y/o told me she half thinks Santa is real and half thinks Santa isn't real, and is this a good age for me to just...you know....tell her...about the Bayesian approach to probability?

23 26 400

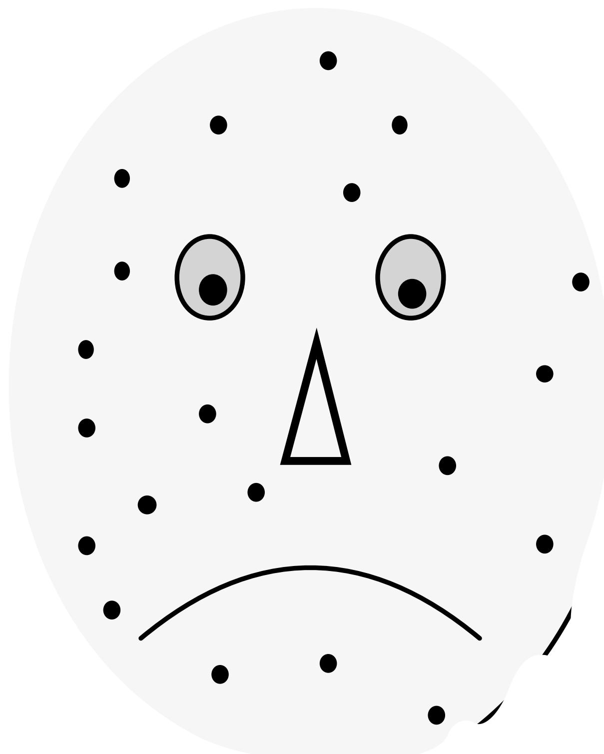
2. Bayes' rule and Bayesian inference

You wake up with spots on your face !

Is it more likely that you have smallpox or chickenpox knowing that you read that there is:

« 90 % of chance that you have spots when you have smallpox »

« 80 % of chance that you have spots when you have chickenpox »



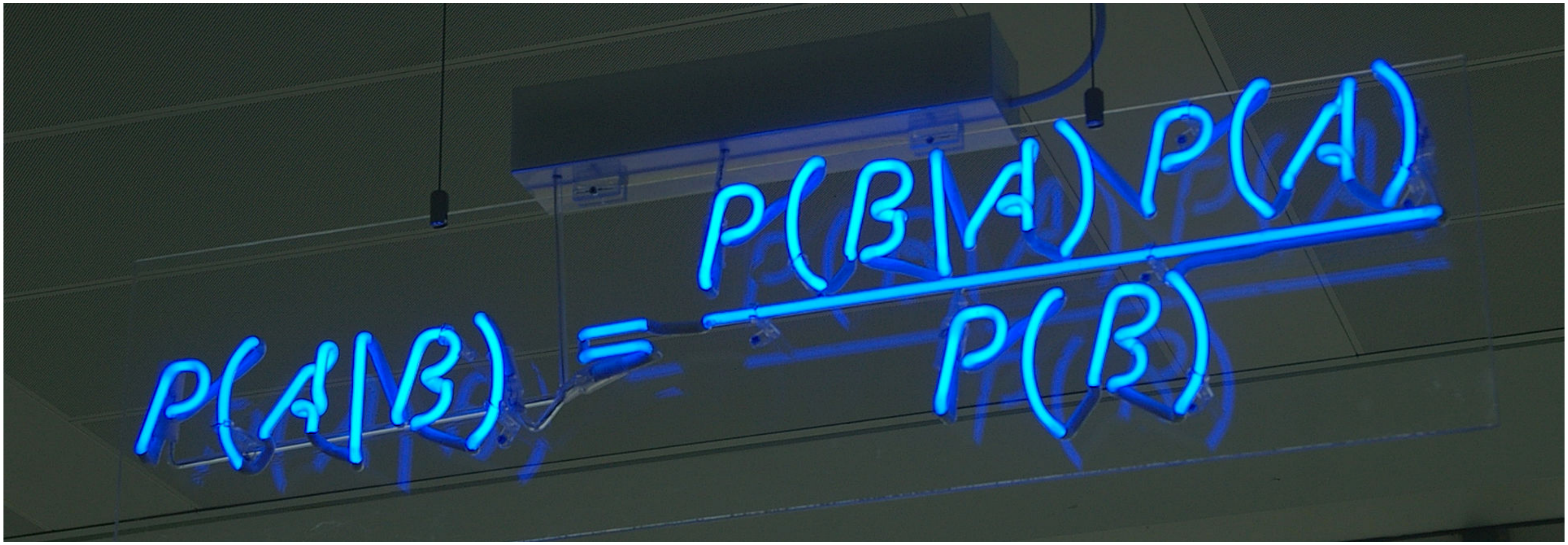
To answer this question you also have to take into account the prevalence of each disease …

Thomas Bayes
(1701-1761)

Pierre-Simon Laplace
(1749-1827)

+ *An Essay towards solving a Problem in the Doctrine of Chances* (1763)
  Published by Richard Price

BAYES' RULE / LAW / THEOREM



$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

BDA : THE FRAMEWORK

How the experience should change our measure of uncertainty on a parameter θ ? It suppose :

-An **a priori** opinion on parameter **θ** before the experiment : P(**θ**)

- Some information from the experience (**likelihood**)

The integration of these 2 informations through **Bayes' theorem** gives us :

- An **a posteriori** confidence in **θ** : **Probability of θ given the data**

$$P(\theta \mid x)$$

Applying this to Bayes' Rule

We want to compute the a posteriori : $P(\theta \mid x)$

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

then A = parameter : θ and B = data : x

BAYESIAN INFERENCE

LIKELIHOOD
How probable is **the data** *given that our parameter is the good one* ?

PRIOR
How probable was our parameter before observing the evidence ?

$$P(\theta \mid x) = \frac{P(x \mid \theta) \,.\, P(\theta)}{P(x)}$$

POSTERIOR
How probable is **our parameter** *given the observed data* ?

MARGINAL
How probable is the new evidence under all possible hypotheses ?

The **probability that the hypothesis is true**, **given the evidence**, is equal to the **likelihood of the evidence occurring when the hypothesis is true**, times the **probability of the hypothesis being true before seeing any evidence**, divided by the **probability of the evidence occurring under all possible hypotheses**.

Bayesian inference

$$\text{Posterior} = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Marginal Likelihood}}$$

BAYESIAN INFERENCE

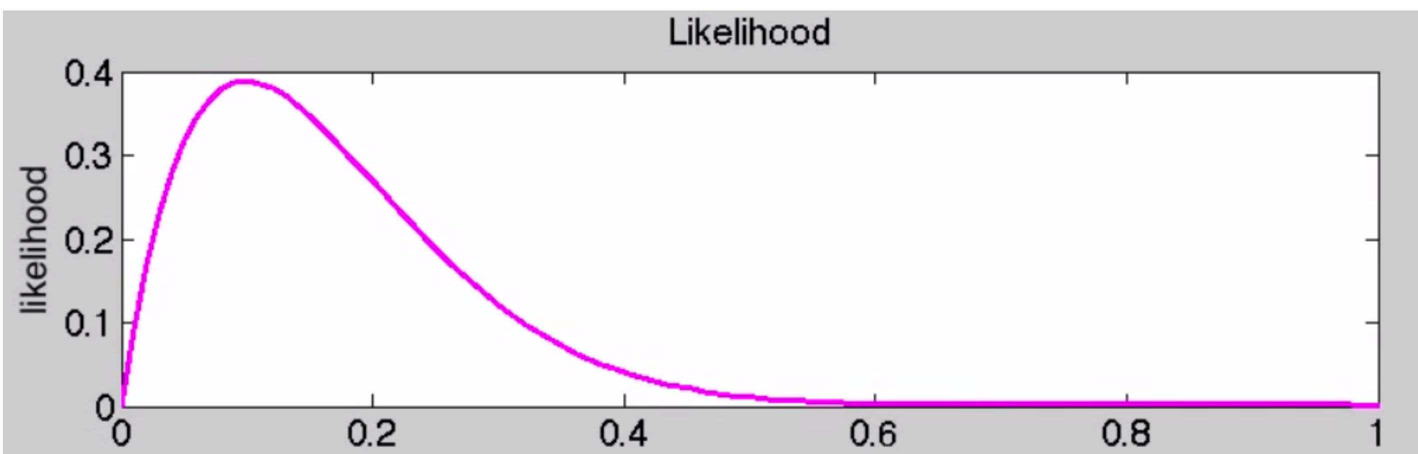
$$\text{Posterior} \propto \text{Likelihood} \cdot \text{Prior}$$

LIKELIHOOD FUNCTION

$P(x \mid \theta) = P(\text{ data} \mid \text{parameter})$

When the data x are known, the likelihood function quantifies **how certain values of the parameter θ are consistent with the observed values x.**

We take 10 people from a population , we get 1 person with the disease.

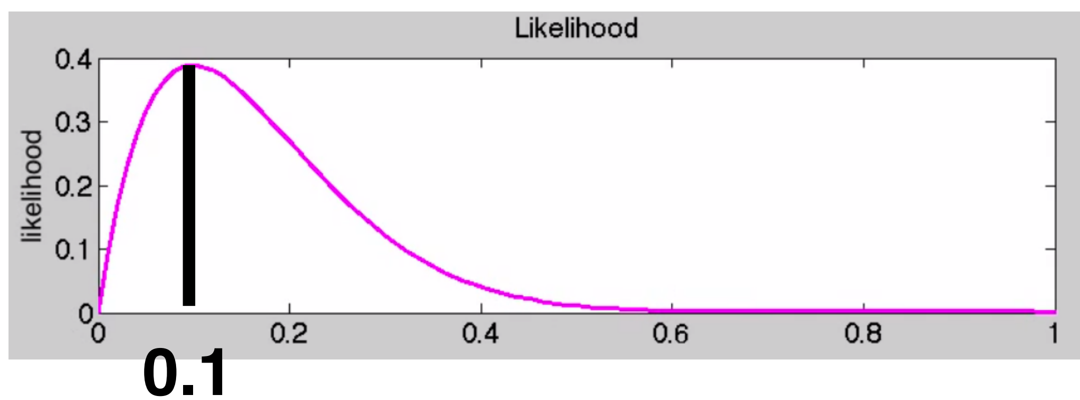


LIKELIHOOD FUNCTION

Frequentist inference can rely on the likelihood function.

—> **Maximum Likelihood Estimation**

a way of estimating statistical parameters by **choosing the parameters that make the data most likely to have happened** **(that maximize the likelihood function)** .



Bayesian inference

$$\text{Posterior} \propto \text{Likelihood} \cdot \text{Prior}$$

PRIOR PROBABILITY DENSITY FUNCTION

Need to have / choose a prior

represent your previous knowledge about the parameter

often seen as subjective (but can be avoided)…

If you know nothing about the parameter, what prior should you choose?

Choose an uninformative (or vague) prior

If you have knowledge about the parameter :

Choose an informative prior

A CONCRETE SITUATION

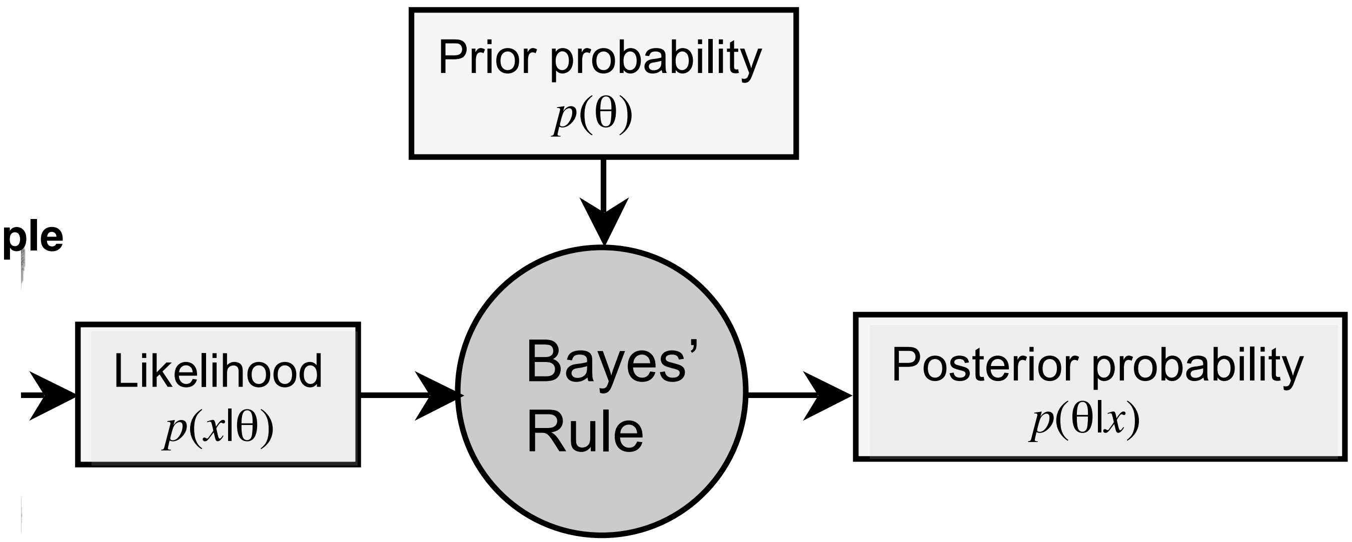
We want to evaluate **the prevalence of a disease in a population** (our theta) from **a sample of N people.** In this sample **X persons appear to have the disease.**

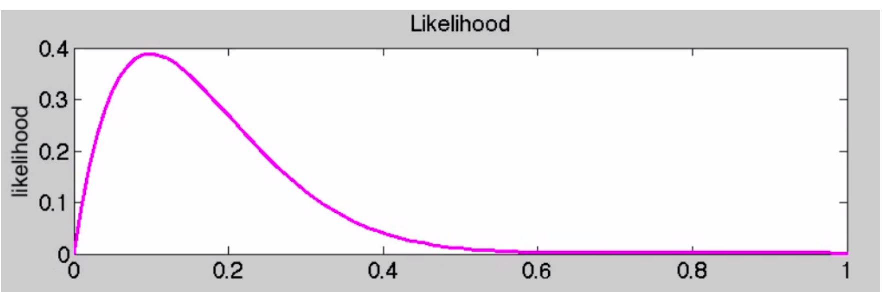
A CONCRETE SITUATION

Informative or not

DATA from sample

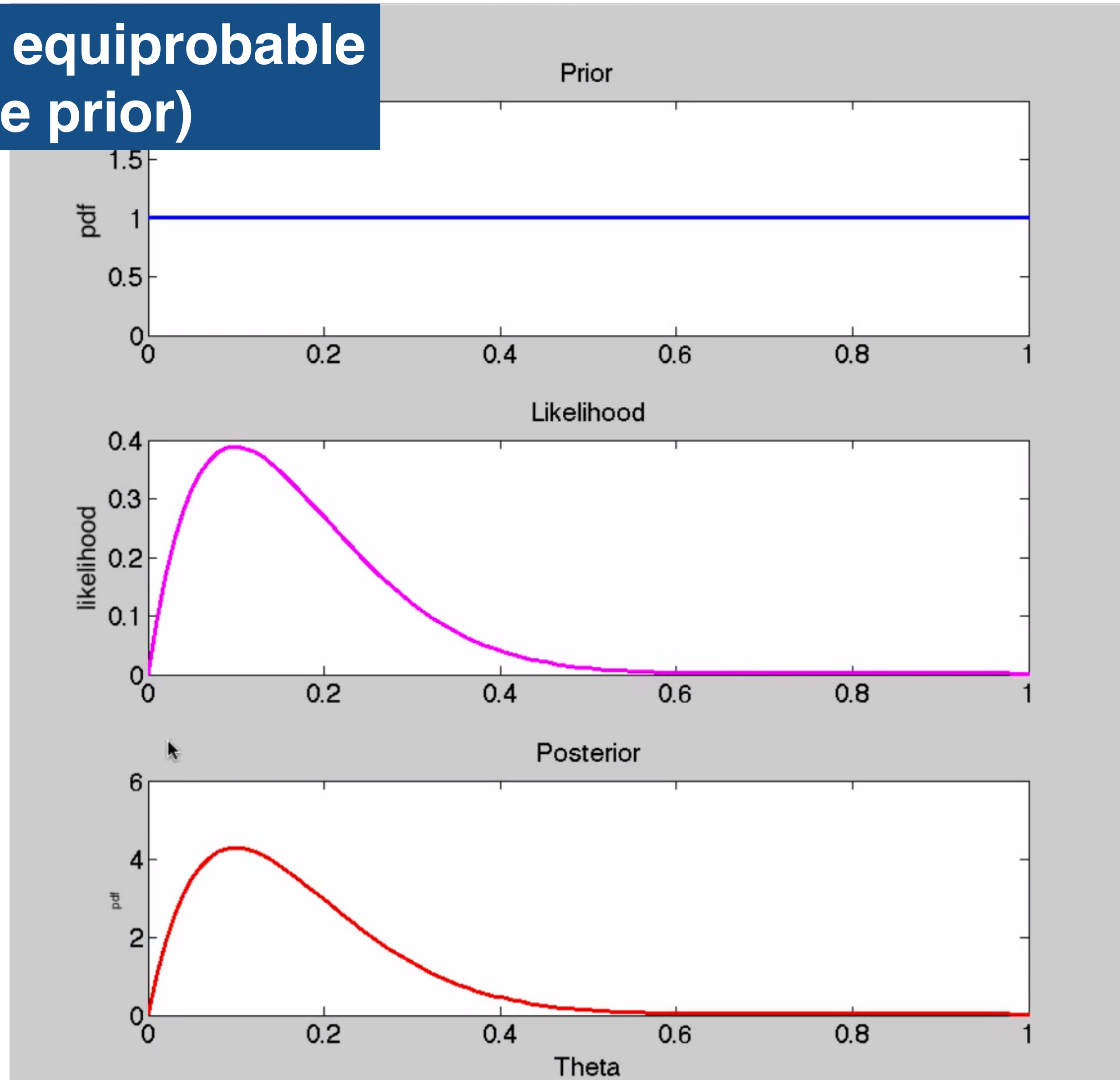
X=1
N=10





Non informative prior with few data

All values of theta are equiprobable (non-informative prior)

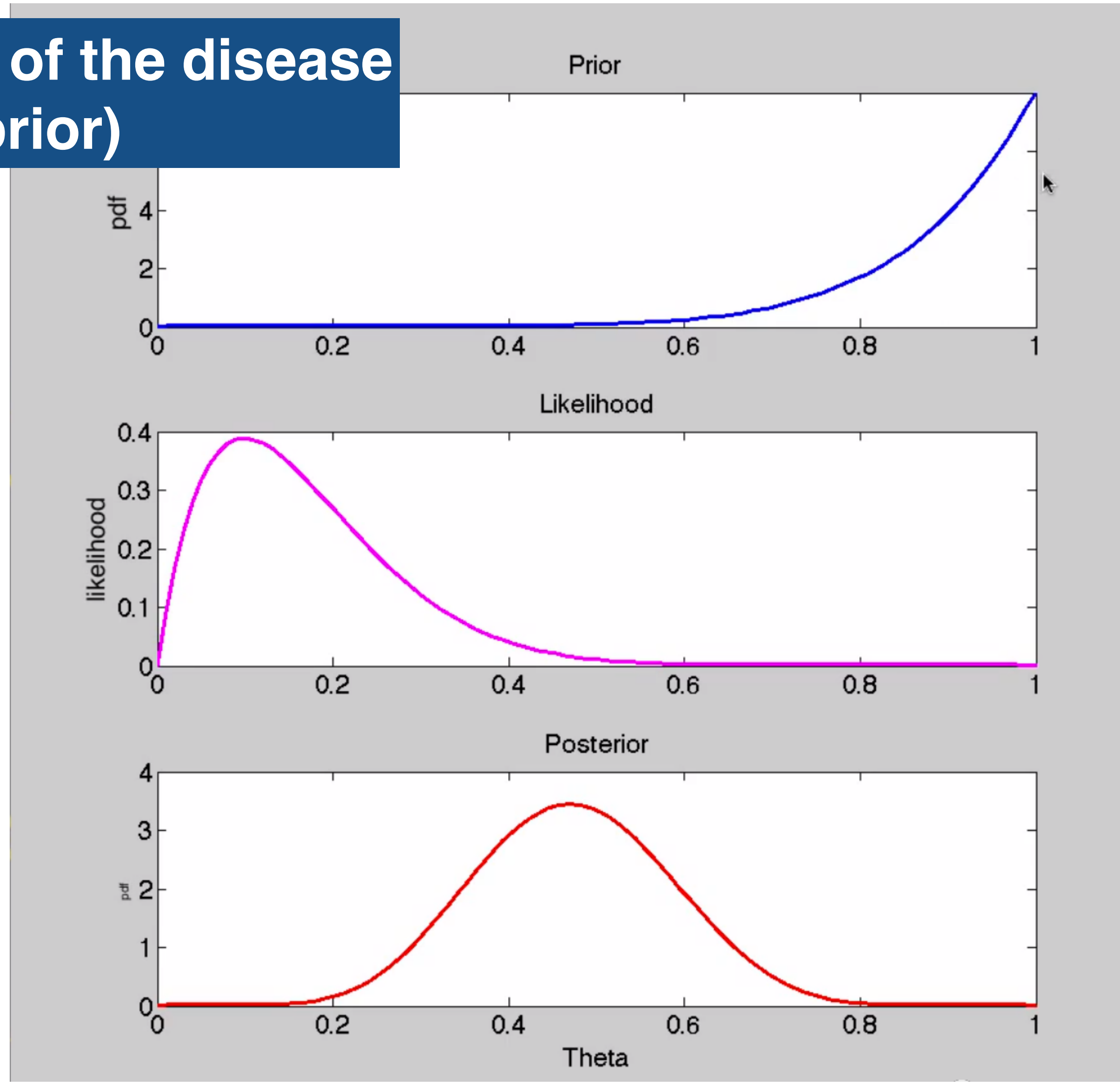


X=1
N=10

The a posteriori distribution will be **close to the likelihood if the prior is non-informative** (or if you have a lot of data)

Informative prior with few data

Prior = high prevalence of the disease (informative prior)



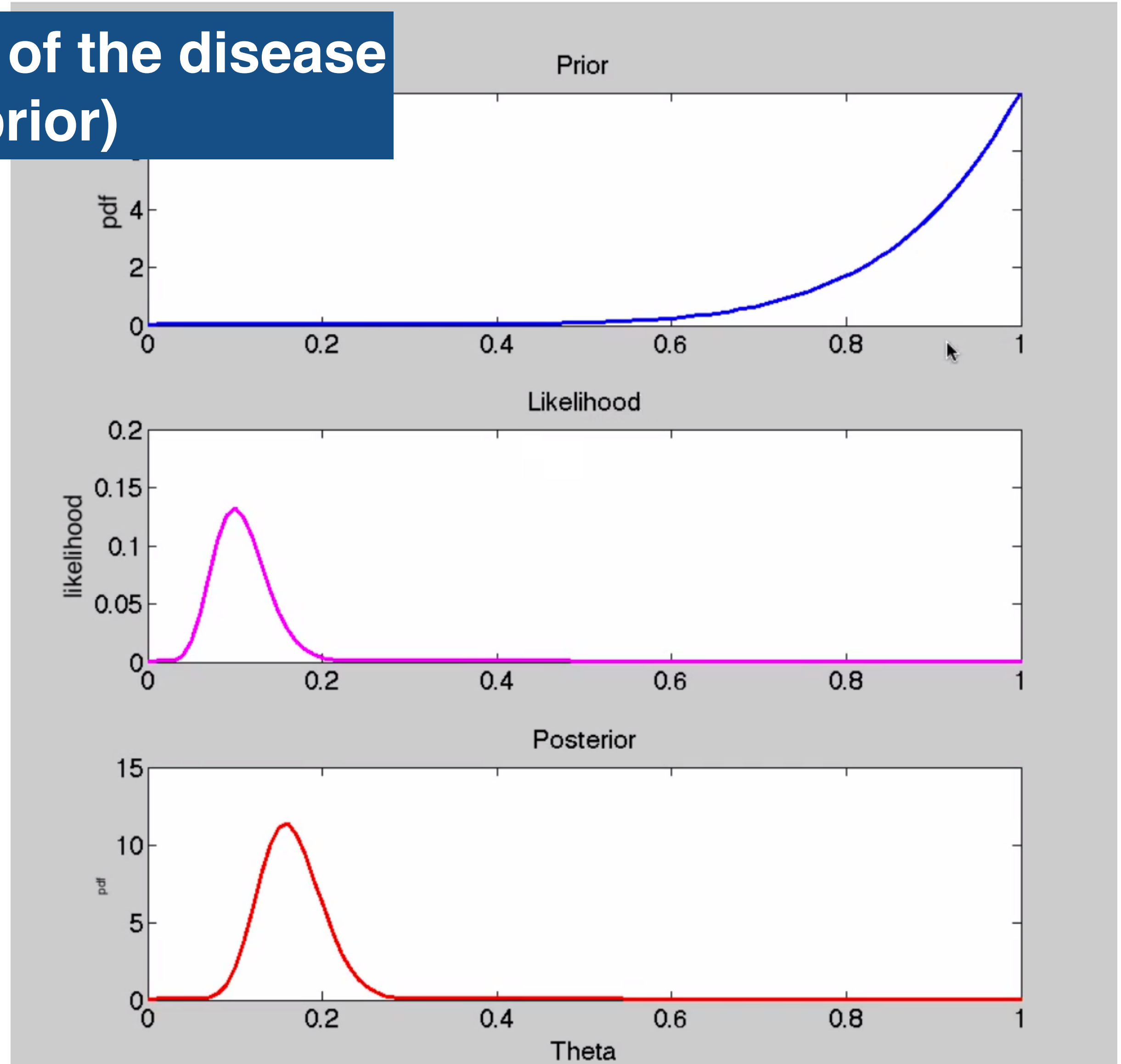
X=1
N=10

The a posteriori distribution will be **close to/influenced by the prior distribution** if it is informative and if there is little data.

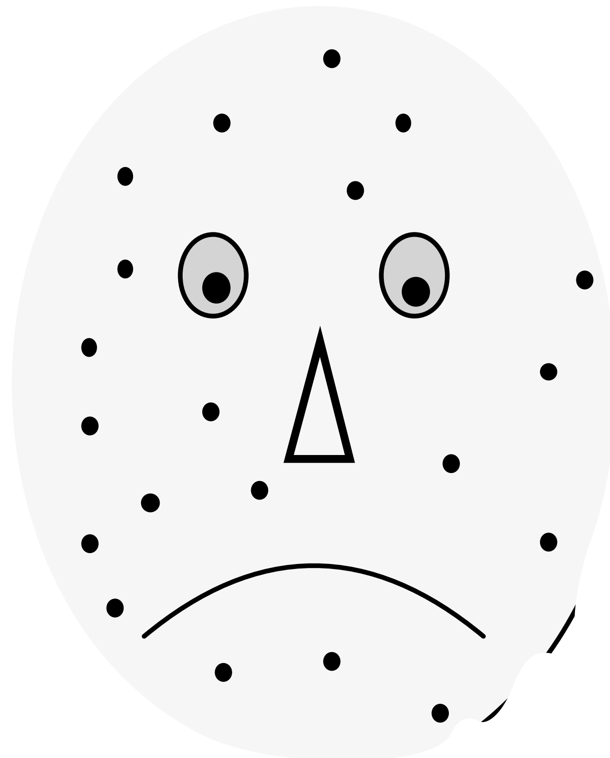
Informative prior with more data

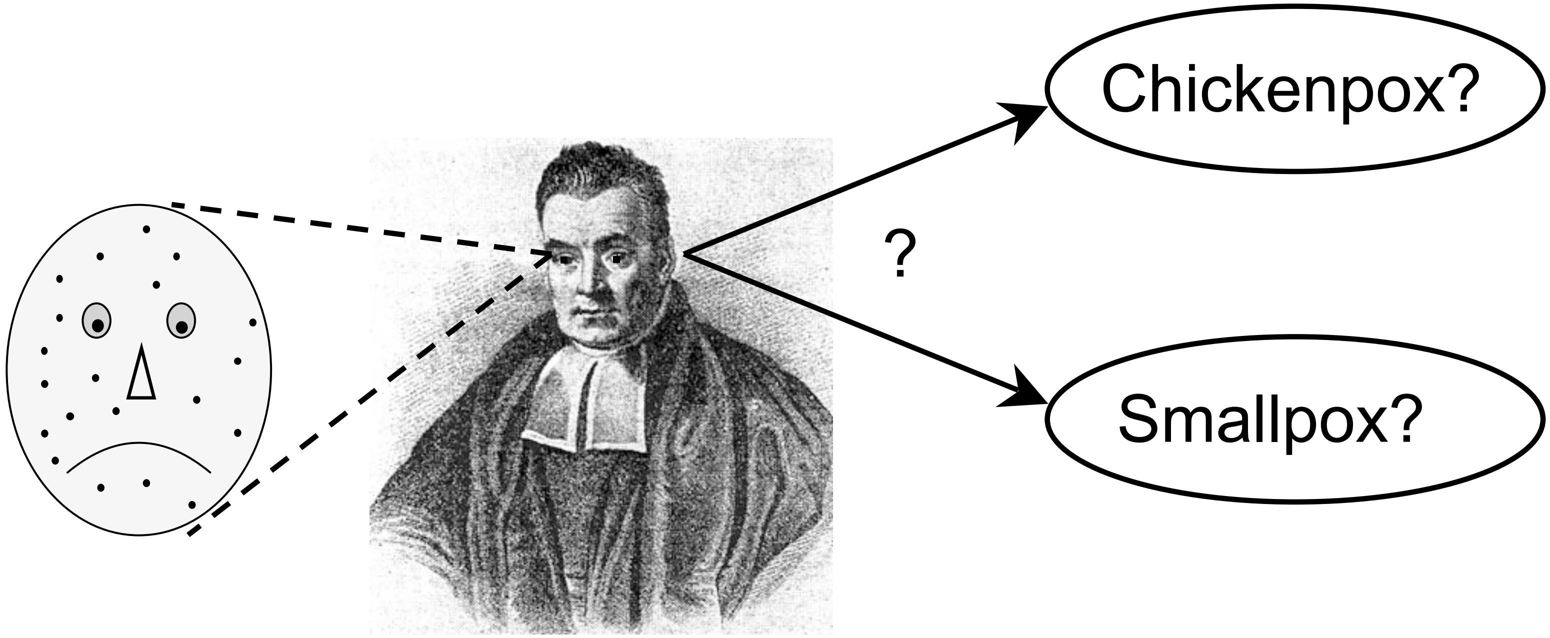
Prior = high prevalence of the disease (informative prior)

X=10
N=100



The more you have data, the more you are confident in your model, **the less you rely on your prior**





θ = parameter value = disease

x = observed data = having spots

« 90 % of chance that you have spots when you have smallpox »

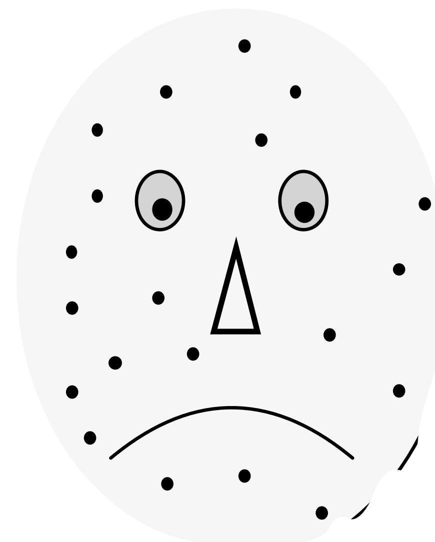
« 80 % of chance that you have spots when you have chickenpox »

Probability to have spots given that you have smallpox

$$P(\text{ spots} \mid \text{smallpox}) = 0.9$$

Probability to have spots given that you have chickenpox

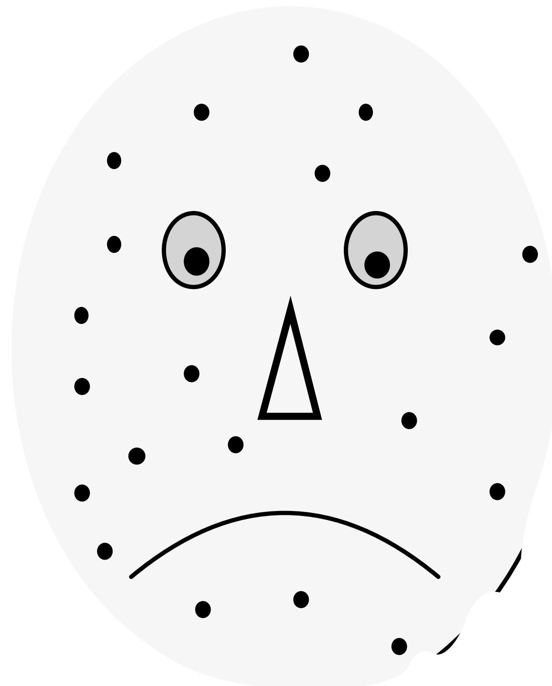
$$P(\text{ spots} \mid \text{chickenpox}) = 0.8$$



$P(x \mid \theta)$ = **likelihood !**

MLE = $P(\text{ spots} \mid \text{smallpox})$

then my best guess is I have smallpox



«% of chance that you have smallpox when you have spots on your face »

P(smallpox | spots)

Probability that you have smallpox given that you have spots on your face

LIKELIHOOD
How probable is **the data** *given that our hypothesis is true ?*

PRIOR
How probable was our hypothesis before observing the evidence ?

$$P(\theta \mid x) = \frac{P(x \mid \theta) \cdot P(\theta)}{P(x)}$$

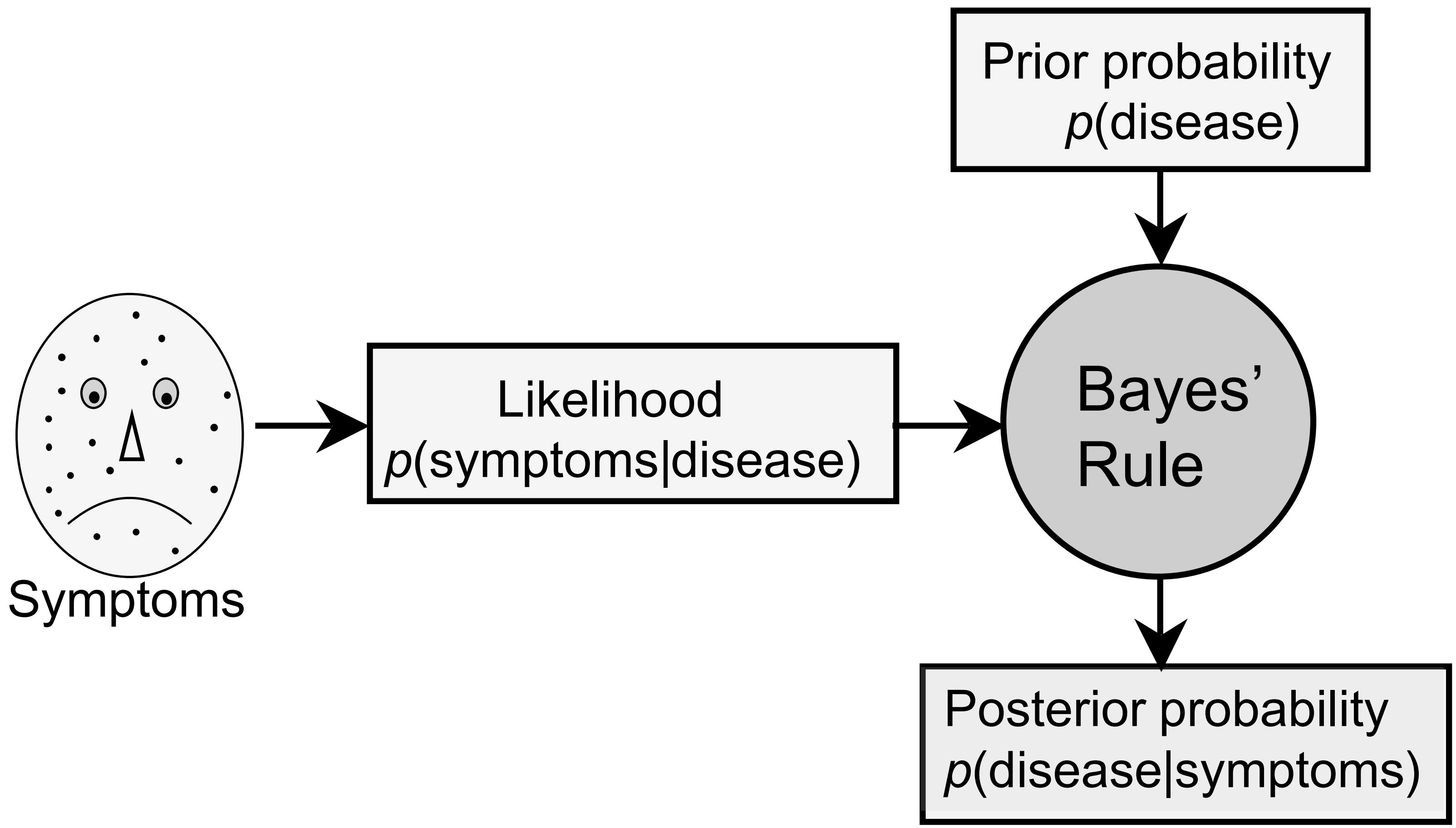
POSTERIOR
How probable is **our hypothesis** *given the observed data ?*

MARGINAL
How probable is the new evidence under all possible hypotheses ?

with θ = chickenpox or smallpox and x = spots presence

chickenpox $=\theta_C$

smallpox $=\theta_S$


Prior probability
p(disease)
Symptoms
Likelihood
p(symptoms|disease)
Bayes' Rule
Posterior probability
p(disease|symptoms)

Bayes' Theorem applied to Smallpox

$$P(\text{smallpox} \mid \text{spots}) = \frac{\overbrace{P(\text{spots} \mid \text{smallpox})}^{=\,0{,}9} \cdot P(\text{smallpox})}{P(\text{spots})}$$

$$P(\text{smallpox}) = 0.001$$

Probability that a randomly chosen individual has smallpox = **prevalence**

$$P(\text{spots}) = 0.081$$

Probability that a randomly chosen individual has spots on his face

Bayes' Theorem applied to Smallpox

$$P(\text{smallpox} \mid \text{spots}) = \frac{P(\text{spots} \mid \text{smallpox}) \cdot P(\text{smallpox})}{P(\text{spots})}$$

$$P(\text{smallpox} \mid \text{spots}) = \frac{0{,}9 \cdot 0.001}{0.081}$$

$$P(\text{smallpox} \mid \text{spots}) = 0.011$$

Bayes' Theorem applied to Chickenpox

$$P(\text{chickenpox} \mid \text{spots}) = \frac{\overset{0{,}8}{P(\text{spots} \mid \text{chickenpox})} \cdot P(\text{chickenpox})}{P(\text{spots})}$$

$$P(\text{chickenpox}) = 0.1$$

Probability that a randomly chosen individual has chickenpox = **prevalence**

$$P(\text{spots}) = 0.081$$

Probability that a randomly chosen individual has spots on his face

Bayes' Theorem applied to Smallpox

$$P(\text{chickenpox} \mid \text{spots}) = \frac{P(\text{spots} \mid \text{chickenpox}) \cdot P(\text{chickenpox})}{P(\text{spots})}$$

$$P(\text{chickenpox} \mid \text{spots}) = \frac{0{,}8 \cdot 0.1}{0.081}$$

$$P(\text{chickenpox} \mid \text{spots}) = 0.988$$

Bayes' Theorem applied to Smallpox

$$P(\text{chickenpox} \mid \text{spots}) = \frac{P(\text{spots} \mid \text{chickenpox}) \cdot P(\text{chickenpox})}{P(\text{spots})}$$

$$P(\text{chickenpox} \mid \text{spots}) = \frac{0{,}8 \cdot 0.1}{0.081}$$

$$P(\text{chickenpox} \mid \text{spots}) = 0.988$$

Posterior Probabilities

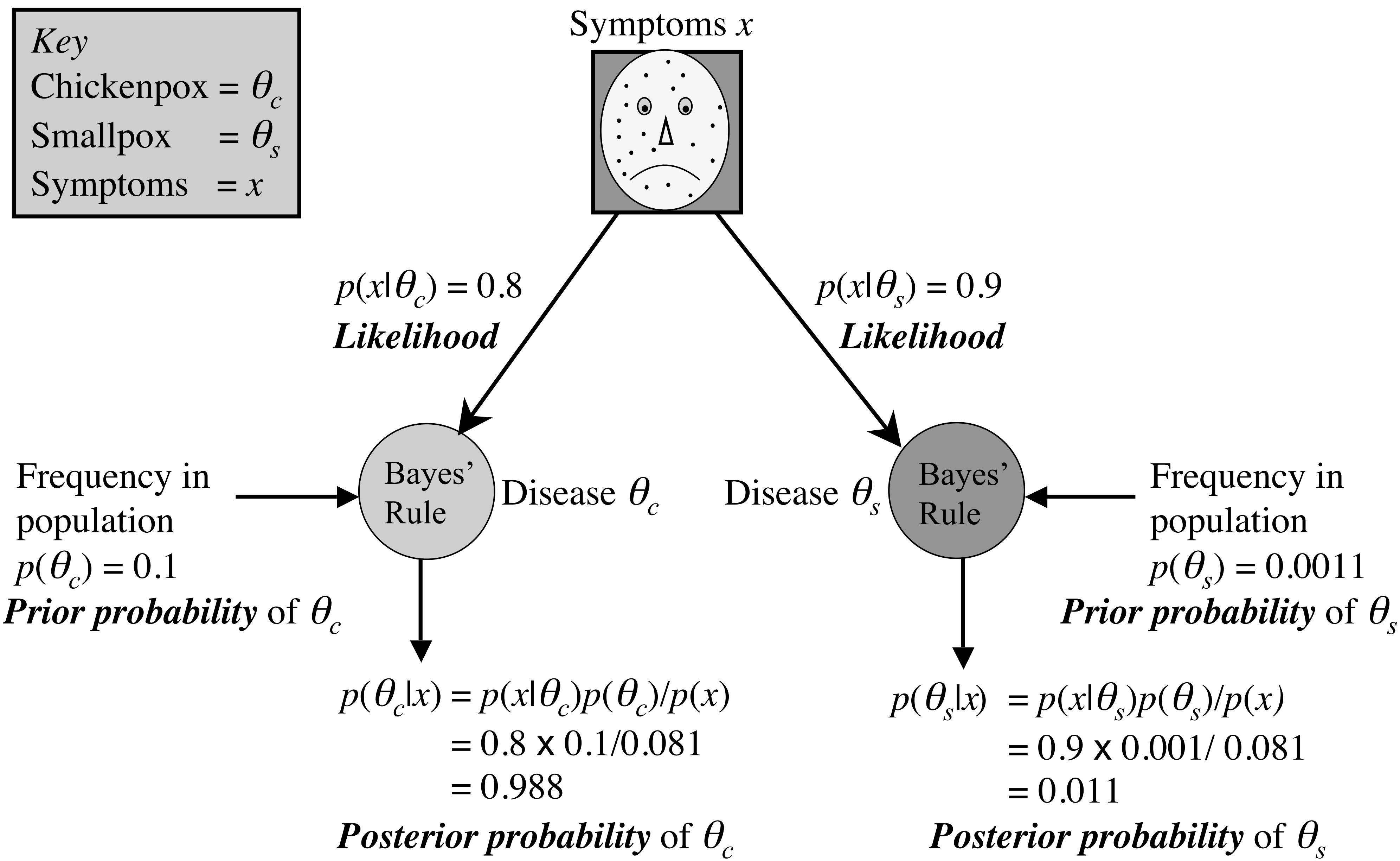
$$P(\text{ smallpox} \mid \text{spots}) = 0.011$$

$$P(\text{ chickenpox} \mid \text{spots}) = 0.988$$

smallpox $=\theta_s$	chickenpox $=\theta_c$	spots $=x$

$$P(\theta_s \mid x) = 0.011$$

$$P(\theta_c \mid x) = 0.988$$



Key
Chickenpox = θ_c
Smallpox = θ_s
Symptoms = x
Symptoms x
$p(x|\theta_c) = 0.8$
Likelihood
$p(x|\theta_s) = 0.9$
Likelihood
Frequency in population
$p(\theta_c) = 0.1$
Prior probability of θ_c
Bayes' Rule
Disease θ_c
Disease θ_s
Bayes' Rule
Frequency in population
$p(\theta_s) = 0.0011$
Prior probability of θ_s
$p(\theta_c|x) = p(x|\theta_c)p(\theta_c)/p(x)$
$= 0.8 \times 0.1/0.081$
$= 0.988$
Posterior probability of θ_c
$p(\theta_s|x) = p(x|\theta_s)p(\theta_s)/p(x)$
$= 0.9 \times 0.001/\ 0.081$
$= 0.011$
Posterior probability of θ_s

MODEL SELECTION

We wish to select the most probable model, this is known as model selection, which involves a comparison using a ratio of posterior probabilities.

$$R_{post} = \frac{P(\theta_c \mid x)}{P(\theta_s \mid x)}$$

$$R_{post} = \frac{P(\theta_c \mid x)}{P(\theta_s \mid x)}$$

$$R_{post} = \frac{(P(x \mid \theta_c) \cdot P(\theta_c)) / P(x)}{(P(x \mid \theta_s) \cdot P(\theta_s)) / P(x)}$$

$$R_{post} = \frac{P(x \mid \theta_c)}{P(x \mid \theta_s)} \cdot \frac{P(\theta_c)}{P(\theta_s)}$$

Ration of likelihood or BAYES FACTOR

ratio of priors or PRIOR ODDS

$$R_{post} = \frac{0.80}{0.90} \cdot \frac{0.1}{0.001}$$

$$R_{post} = 88.9$$

If the posterior odds is greater than 3 or less than 1/3 (in both cases one hypothesis is more than 3 times more probable than the other) then this is considered to represent a substantial difference between the probabilities of the two hypotheses

3. Bayesian vs. Frequentist Inference

INFERENCE

PROBLEMS WITH NHST

- p-values misinterpretations :: p(data | H_0)
- Uses p(data | H_0) to make a decision regarding p(H_0| data) using *modus tollens* syllogism that does not seem to work with uncertain statements ("very likely that")

- frequentist inference depends on the subjective intentions of the researcher :: p-values are dependent on unobserved data and decisions that were never made

- statistically significant result can always be obtained
- Does not take into account previous results

RESULTING FROM INFERENCE

Frequentist Inference

A **unique value** , a « best fit » for which parameter make the data most likely to happen (Maximum Likelihood Estimator)

Confidence intervals around the best fit.

But the parameters can not and should not be probabilized

Bayesian Inference

The parameters may be probabilized.

A full probability distribution of the parameter that reflect our uncertainty

DATA/ PARAMETER - FIXED / RANDOM

The **Frequentist framework** is somewhat **counter-intuitive**

data is assumed to be **random**, **parameter** assumed to be **fixed**

(—> Maximum Likelihood Estimator depends on **P(x | θ)**)

When drawing inference within a **Bayesian framework**

the **data** are treated as a **fixed quantity**

and the **parameters** are treated as **random variables**.

P(θ | x)

PROBABILITY OVER PARAMETER VALUES

In **Frequentist framework**, only one value for the parameter (the best fit) and one confidence interval. Then **we cannot ask** (or even answer!) **questions such as** :

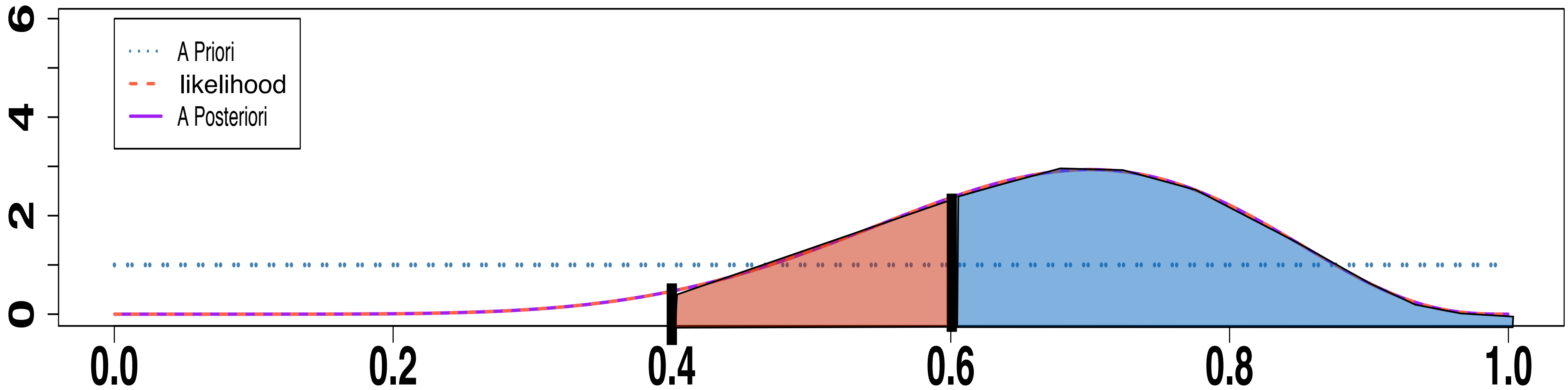
What is the chance that the unknown parameter is greater than 0.4? comprised between 0.4 and 0.6 ?

In **Bayesian framework**, we can !

Non informative prior

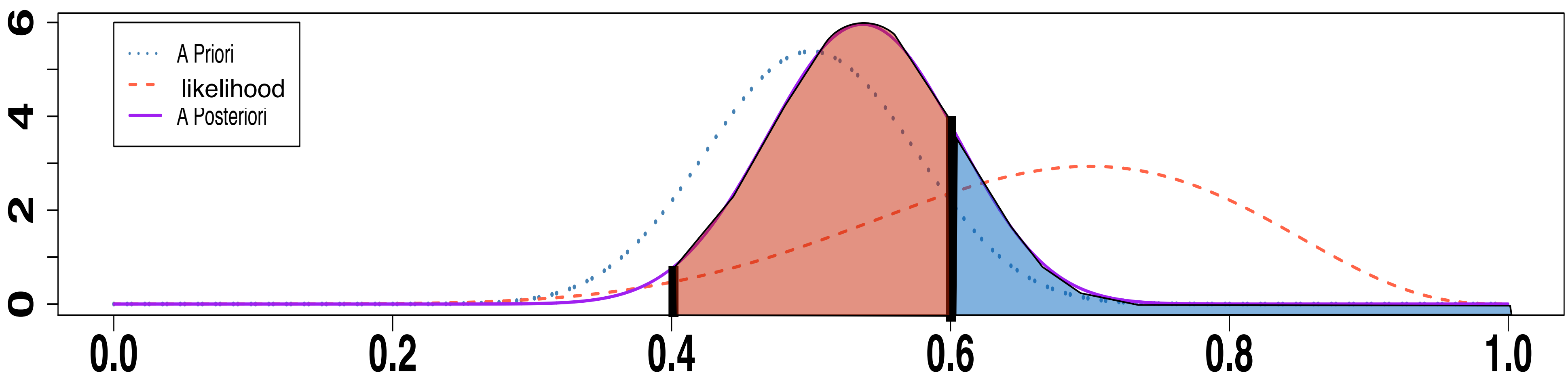
P (θ > 0.6Idata) P(0.4 < θ < 0.6 | data)

All values of theta are equiprobable (a priori non-informative)



Informative prior

There is a 50% chance that theta is between 45% and 55% (subjective a priori).



PREVIOUS KNOWLEDGE

Frequentist framework doesn't take in account **previous knowledge** about the parameters

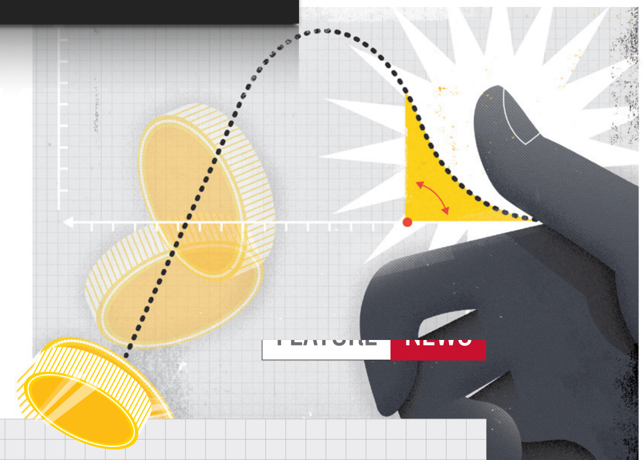
Bayesian framework does

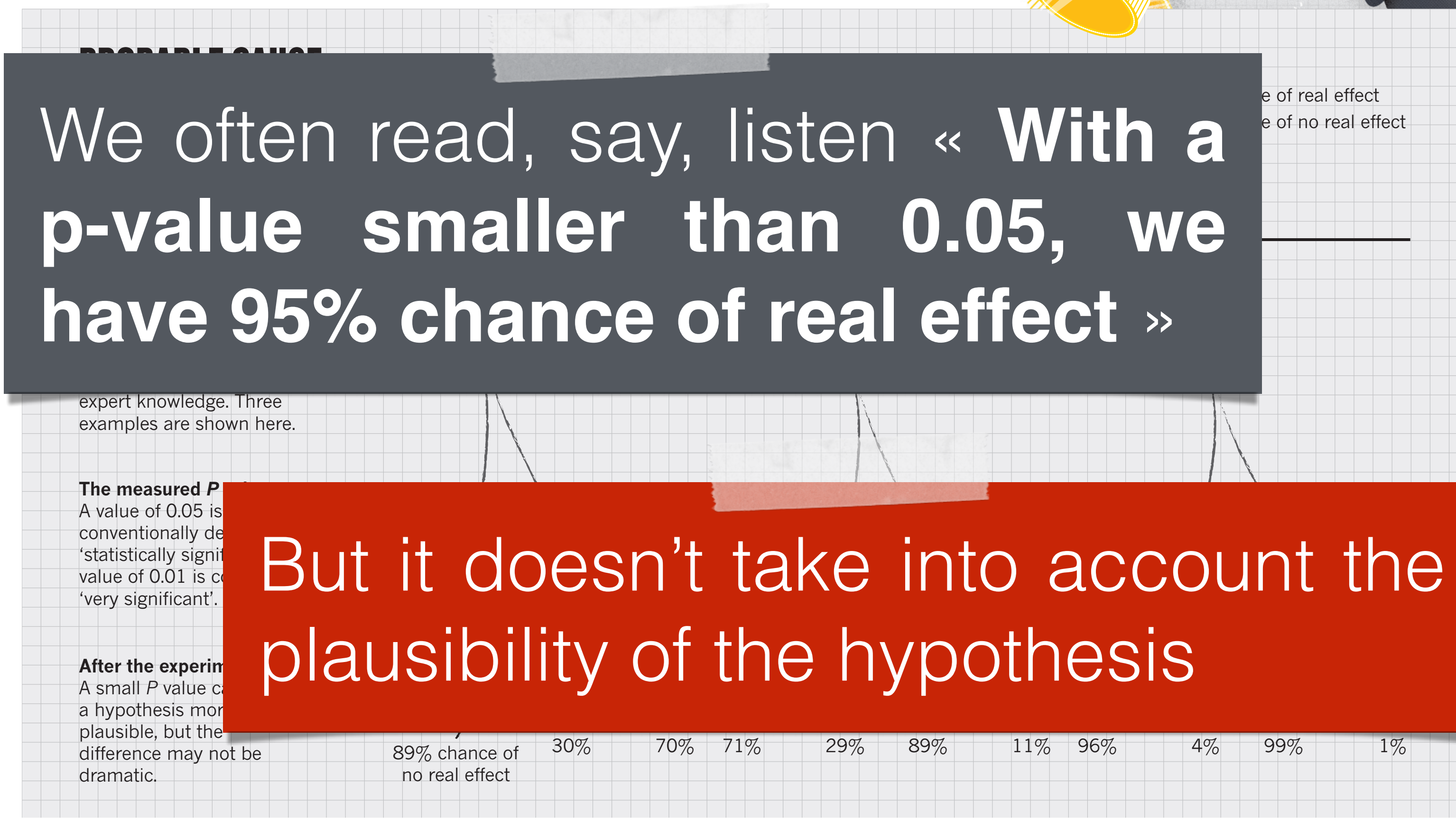
Bayesian vs. frequentist inference

Against p-value as a gold standard

STATISTICAL ERRORS

P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.





We often read, say, listen « **With a p-value smaller than 0.05, we have 95% chance of real effect** »

But it doesn't take into account the plausibility of the hypothesis

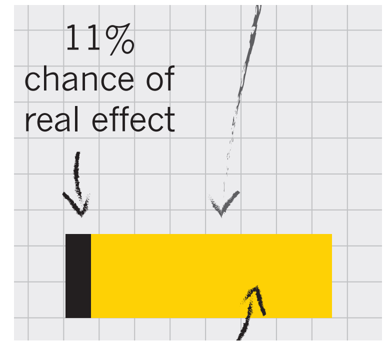
Bayesian vs. frequentist inference

Initial plausibility of the hypothesis before the experiment

Chances of real effect after the experiment

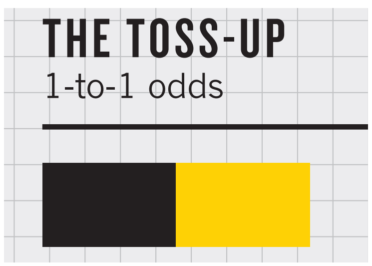
5%

p-value =0.05

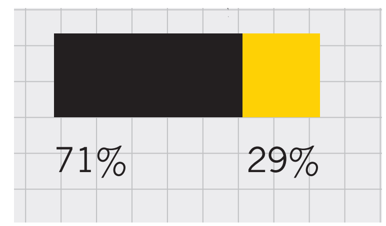


11%

50%

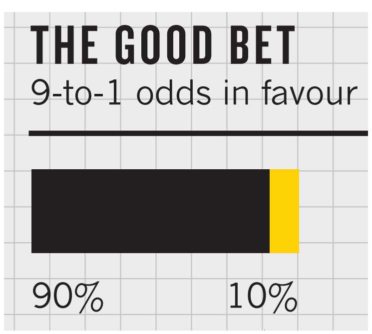


p-value =0.05

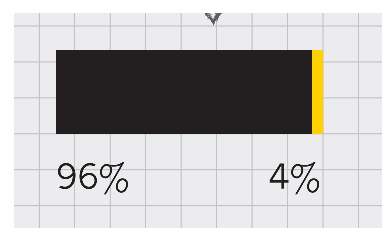


71%

90%



p-value =0.05



96%

BDA IN FEW WORDS

Bayesian Data Analysis is an approach of statistics which

differs in fundamental ways **from frequentist statistics**

is based on **Bayesian inference** ie. applies **Bayes' rule** to solve inferential question of interest

In **Bayesian Data Analysis :**

probability is interpreted as a **degree of belief** rather than the limit frequency of a phenomenon

CONCLUSION

Everything is assigned distributions (prior, posterior)

we are allowed to **incorporate prior information about the parameter** . . .

which is then **updated by using the likelihood function** . . .

leading to the posterior distribution which tell us everything we need about the parameter and allows us to make our inferences

CONCLUSION

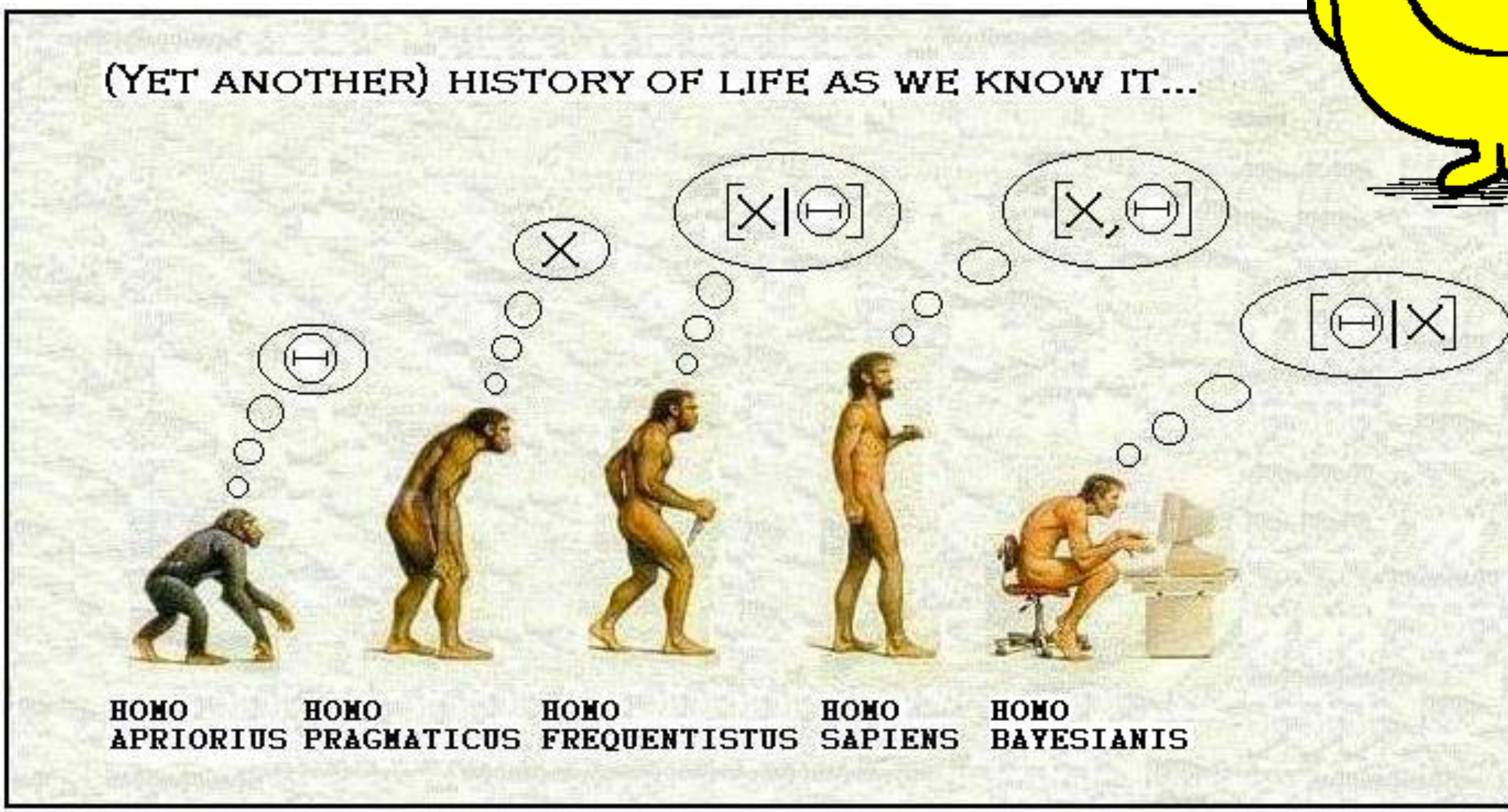
As **parameters** are treated as **random variables,** Bayesian approach offers a **natural framework** to deal with parameter and model uncertainty.

It offers **much more than a single "best fit"** or any sort "sensitivity analysis".

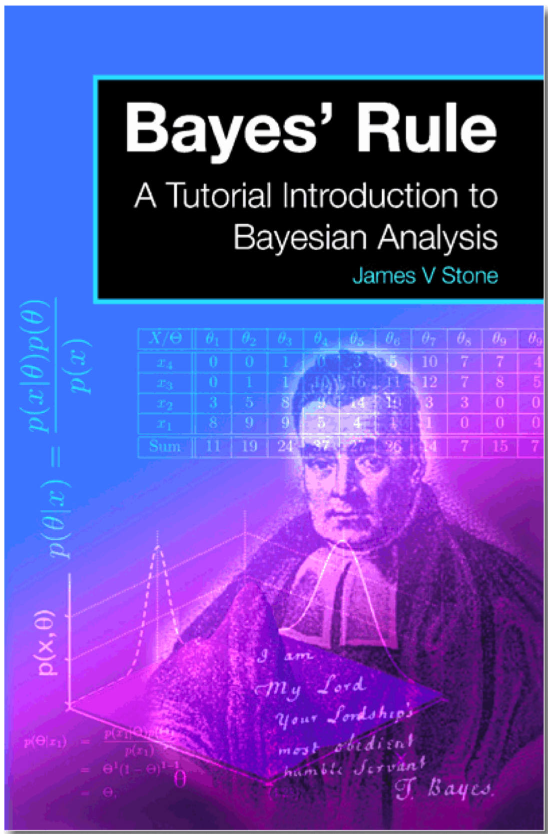
This approach has substantial appeal, and it has become an important part of mainstream statistics.

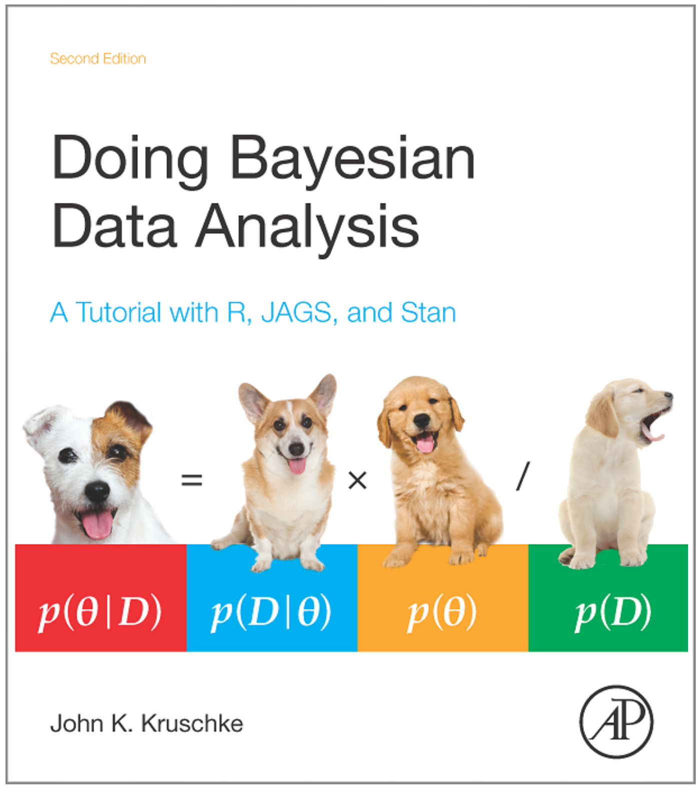
"... we balance probabilities and choose the most likely. It is the scientific use of the imagination ... "

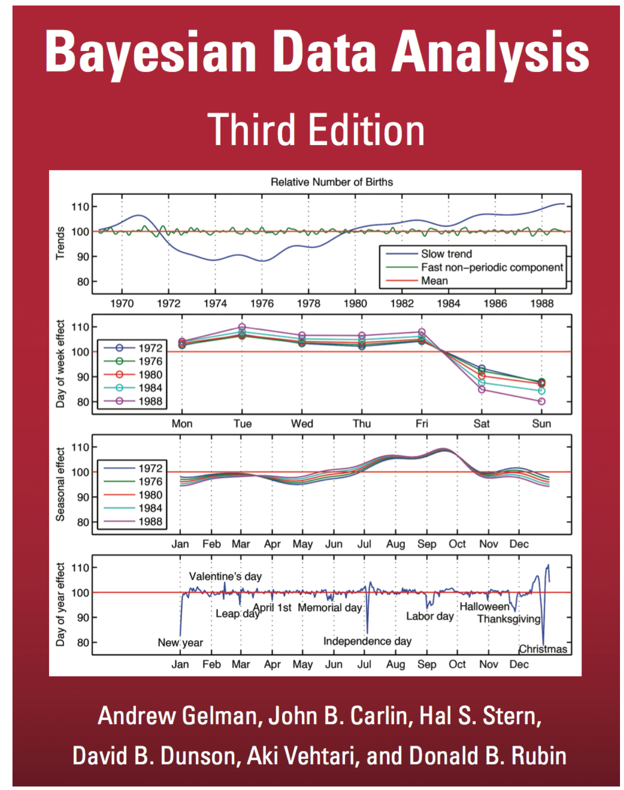
Sherlock Holmes in The Hound of the Baskervilles.
AC Doyle, 1901.


(YET ANOTHER) HISTORY OF LIFE AS WE KNOW IT...
Θ
X
$[X|\Theta]$
$[X,\Theta]$
$[\Theta|X]$
HOMO APRIORIUS
HOMO PRAGMATICUS
HOMO FREQUENTISTUS
HOMO SAPIENS
HOMO BAYESIANIS

SOME REFERENCES







SOME REFERENCES

Stone, J. V. (2013). Bayes' Rule : A tutorial introduction to Bayesian Analysis. Sebetel Press.

Kruschke, J. K. (2010). Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Academic Press / Elsevier.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2002), Bayesian Data Analysis, 2nd Ed., Chap- man & Hall Ltd (London; New York)

Nuzzo, R. (2014). Scientific method: statistical errors. Nature, 506(7487), 150–2. doi:10.1038/506150a

Ox Educ. (2014, August 11). Bayesian statistics : a comprehensive course [Video file]. Retrieved from http://www.youtube.com/watch?v=U1HbB0ATZ_A&index=1&list=PLFDbGp5YzjqXQ4oE4w9GVWdiokWB9gEpm

Niemi J. (2014, January 14). Bayesian statistics [Video file]. Retrieved from http://www.youtube.com/watch?v=Vd_gKry3h4s