BAYES and FREQUENTISM: The Return of an Old Controversy

Louis Lyons

Imperial College and Oxford University

CERN Summer Students July 2015

It is possible to spend a lifetime analysing data without realising that there are two very different fundamental approaches to statistics:

Bayesianism and Frequentism.

How can textbooks not even mention Bayes / Frequentism?

For simplest case $(m \pm \sigma) \leftarrow Gaussian$

with no constraint on $m(true)$ then

$$m - k\sigma < m(true) < m + k\sigma$$

at some probability, for both Bayes and Frequentist

(but different interpretations)

See Bob Cousins "Why isn't every physicist a Bayesian?" Amer Jrnl Phys 63(1995)398

We need to make a statement about

Parameters, Given Data

The basic difference between the two:

Bayesian : Probability (parameter, given data)
(an anathema to a Frequentist!)

Frequentist : Probability (data, given parameter)
(a likelihood function)

PROBABILITY

MATHEMATICAL

Formal

Based on Axioms

FREQUENTIST

Ratio of frequencies as n→ infinity

Repeated “identical” trials

Not applicable to **single event** or **physical constant**

BAYESIAN

Degree of belief

Can be applied to single event or physical constant

(even though these have unique truth)

Varies from person to person ***

Quantified by “fair bet”

Bayesian versus Classical

Bayesian

P(A and B) = P(A;B) x P(B) = P(B;A) x P(A)

e.g. A = event contains t quark

B = event contains W boson

or A = I am in CERN

B = I am giving a lecture

P(A;B) = P(B;A) x P(A) /P(B)

Completely uncontroversial, provided….

Bayesian

$$P(A;B) = \frac{P(B;A) \text{ x } P(A)}{P(B)}$$

Bayes' Theorem

p(param | data) α p(data | param) * p(param)

↑ Posterior ↑ Likelihood ↑ Prior

Problems:

1) p(param) Has particular value

For Bayesian, "Degree of my belief"

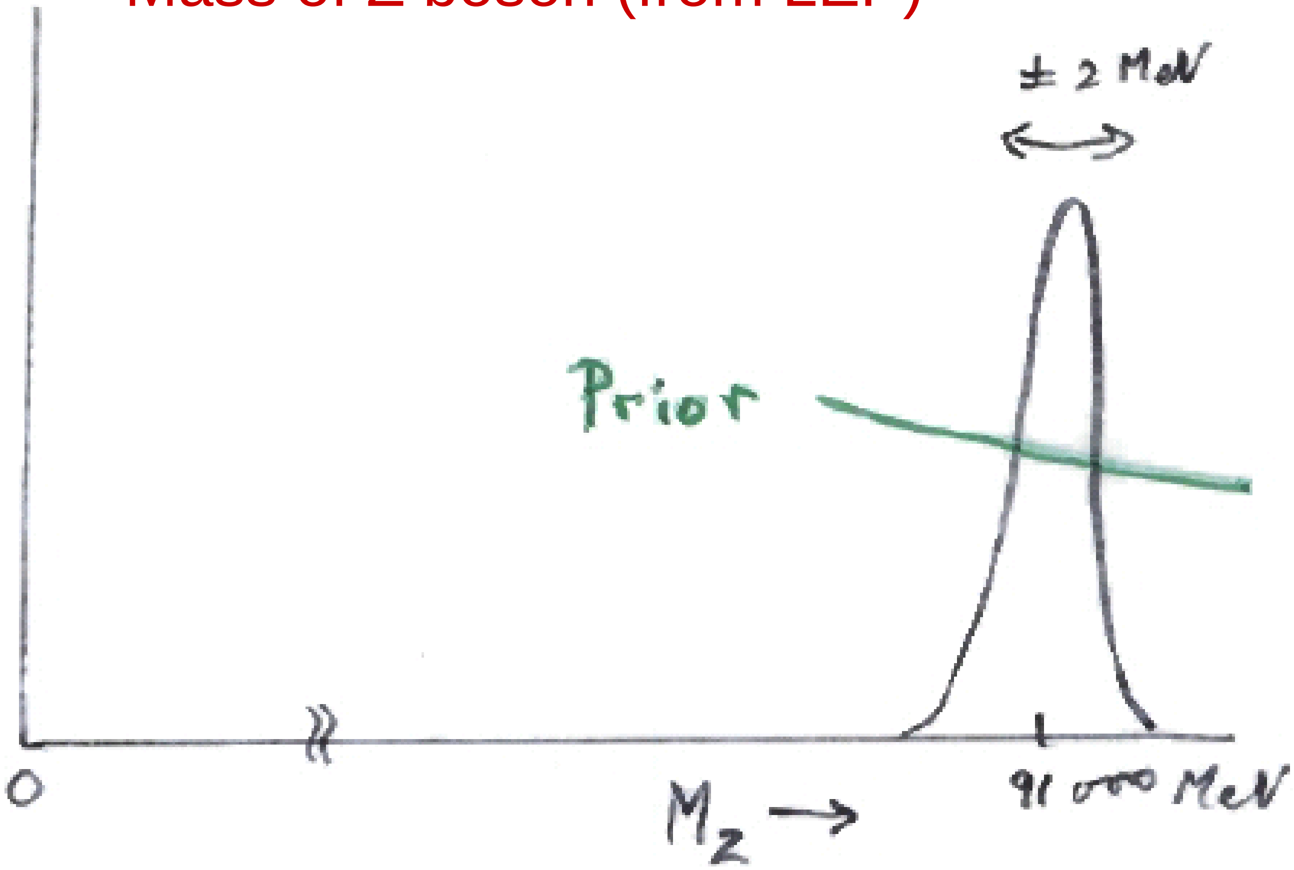
2) Prior What functional form?

Maybe OK if previous measurement

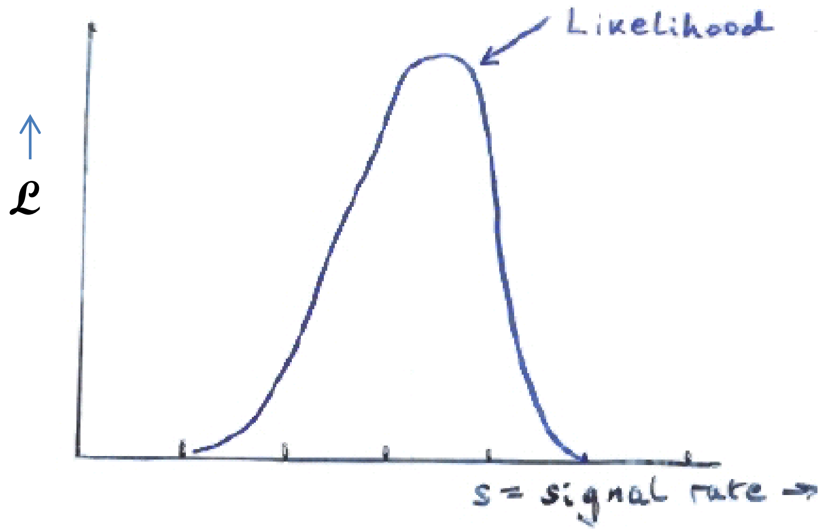
More difficult to parametrise ignorance

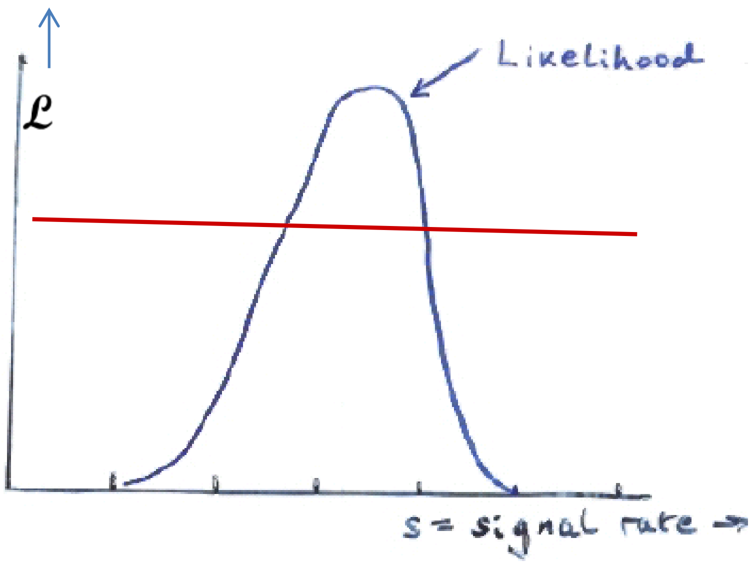
More troubles in many dimensions

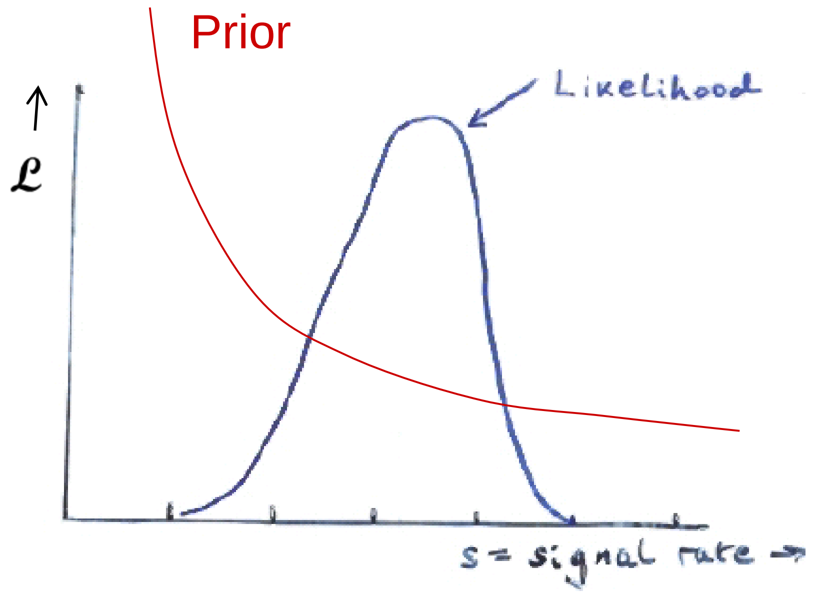
Mass of Z boson (from LEP)



"Data overshadows prior"

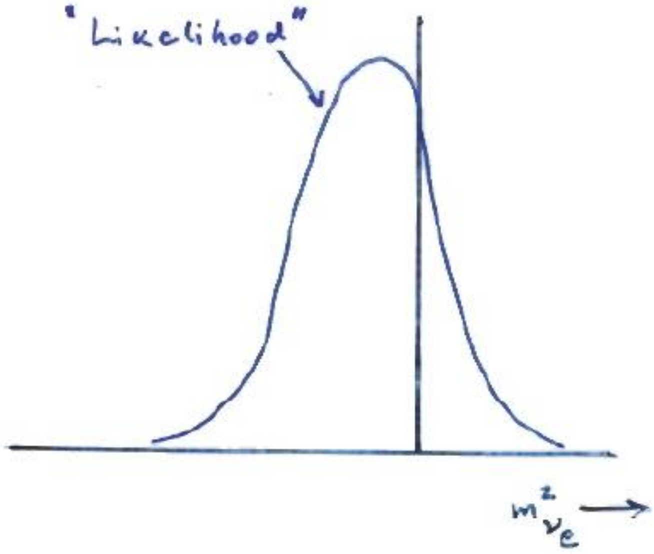


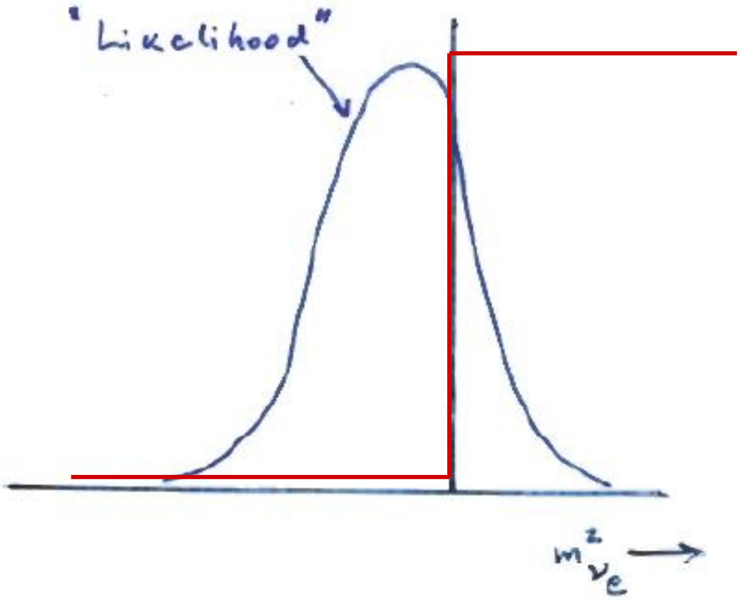




Even more important for UPPER LIMITS

Mass-squared of neutrino



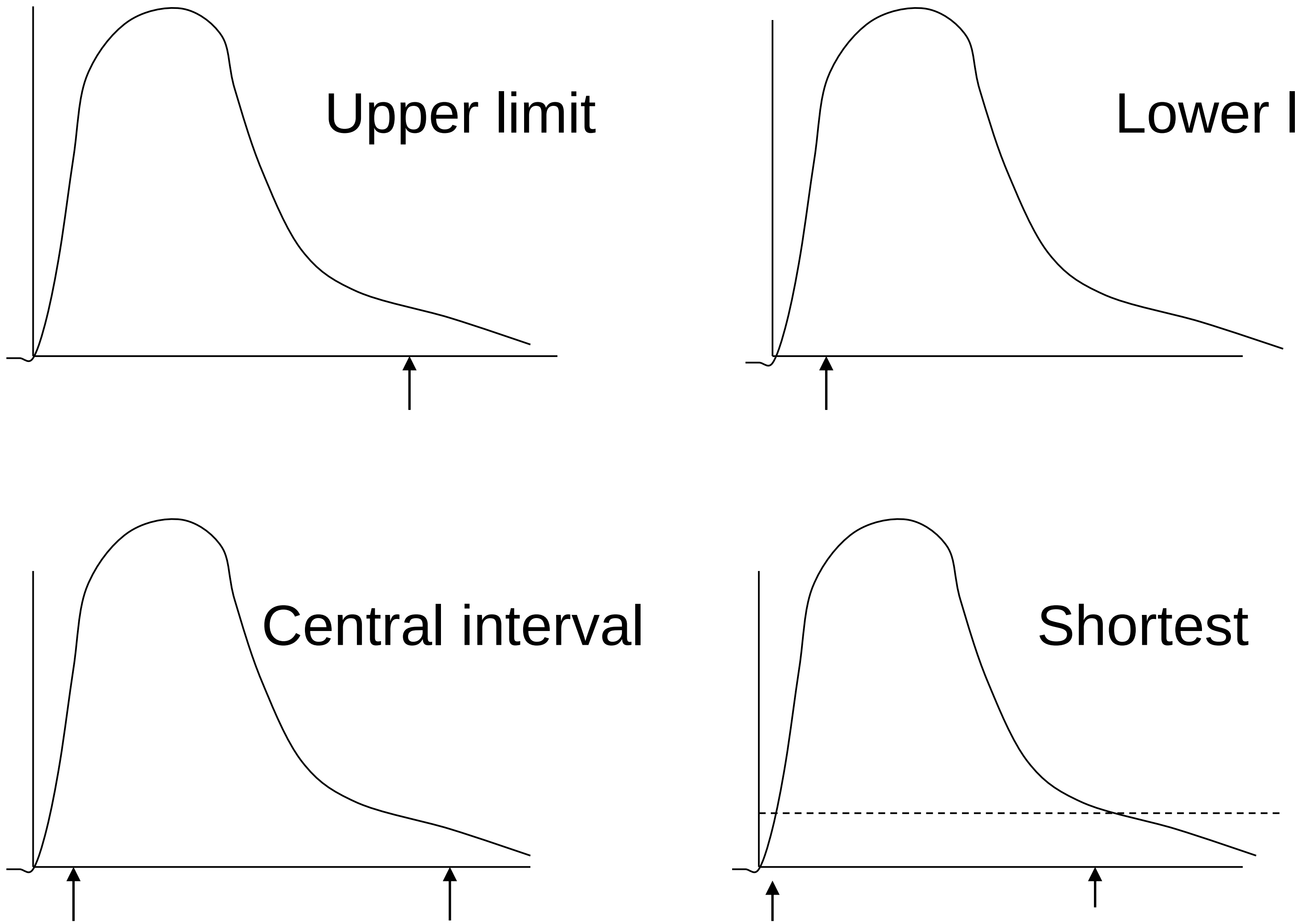


Prior = zero in unphysical region

Posterior for $m^2_{\nu e} = \mathcal{L} \times$ Prior

Bayesian posterior → intervals

(Posterior prob density v parameter)



Example: Is coin fair ?

Toss coin: 5 consecutive tails

What is P(unbiased; data) ? i.e. $p = ½$

Depends on Prior(p)

If village priest: prior ~ $\delta(p = 1/2)$

If stranger in pub: prior ~ 1 for $0 < p < 1$

(also needs cost function)

$$P\ (\text{Data;Theory}) \neq P\ (\text{Theory;Data})$$

P (Data;Theory) $\neq$ P (Theory;Data)

Theory = male or female

Data = pregnant or not pregnant

P (pregnant ; female) ~ 3%

$$P\,(\text{Data};\text{Theory}) \neq P\,(\text{Theory};\text{Data})$$

Theory = male or female

Data = pregnant or not pregnant

P (pregnant ; female) ~ 3%

but

P (female ; pregnant) >>>3%

P (Data;Theory) $\neq$ P (Theory;Data)

HIGGS SEARCH at CERN

Is data consistent with Standard Model?

or with Standard Model + Higgs?

End of Sept 2000: Data not very consistent with S.M.
Prob (Data ; S.M.) < 1% valid frequentist statement

Turned by the press into: Prob (S.M. ; Data) < 1%
and therefore Prob (Higgs ; Data) > 99%

i.e. "It is almost certain that the Higgs has been seen"

Classical Approach

Neyman "confidence interval" avoids pdf for μ

Uses only P(x; μ)

Confidence interval $\mu_1 \rightarrow \mu_2$:

P($\mu_1 \rightarrow \mu_2$ contains μ) = α True for any μ

Varying intervals from ensemble of experiments

fixed

Gives range of μ for which observed value x_0 was "likely" (α)

Contrast Bayes : Degree of belief = α that μ_t is in $\mu_1 \rightarrow \mu_2$

Classical (Neyman) Confidence Intervals

Uses only P(data|theory)

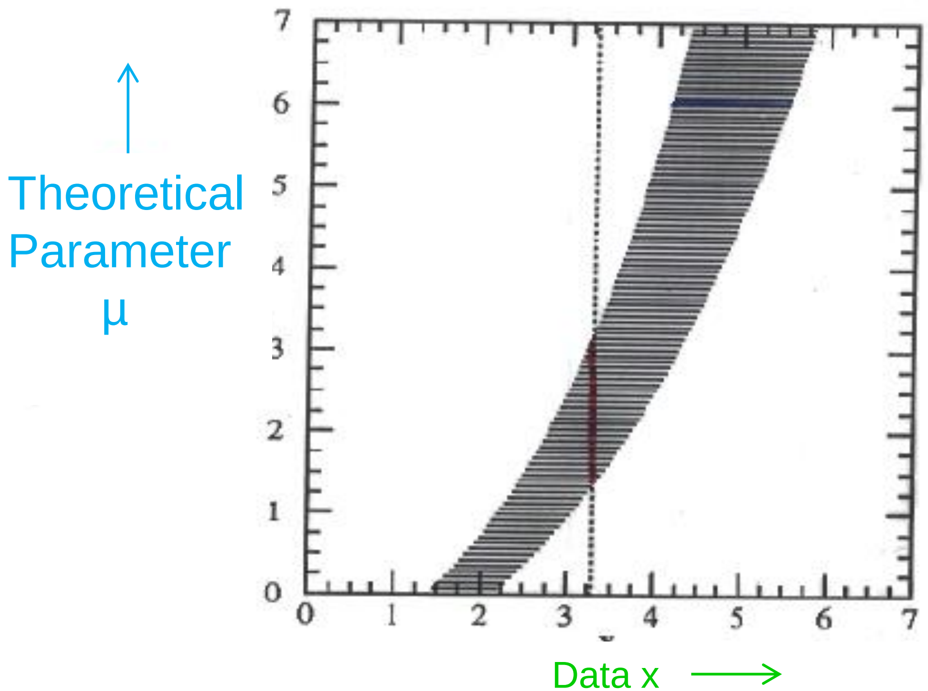


FIG. 1. A generic confidence belt construction and its use. For each value of μ, one draws a horizontal acceptance interval $[x_1, x_2]$ such that $P(x \in [x_1, x_2] | \mu) = \alpha$. Upon performing an experiment to measure x and obtaining the value x_0, one draws the dashed vertical line through x_0. The confidence interval $[\mu_1, \mu_2]$ is the union of all values of μ for which the corresponding acceptance interval is intercepted by the vertical line.

Example:

Param = Temp at centre of Sun

Data = Est. flux of solar neutrinos

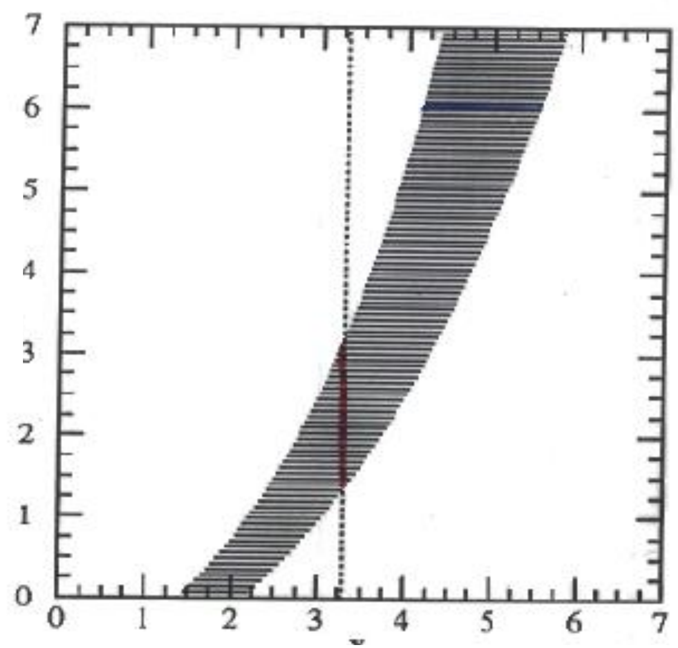
$\mu \geq 0$

No prior for μ

Classical (Neyman) Confidence Intervals

Uses only P(data|theory)

Theoretical Parameter μ



Data x

FIG. 1. A generic confidence belt construction and its use. For each value of μ, one draws a horizontal acceptance interval $[x_1, x_2]$ such that $P(x \in [x_1, x_2] | \mu) = \alpha$. Upon performing an experiment to measure x and obtaining the value x_0, one draws the dashed vertical line through x_0. The confidence interval $[\mu_1, \mu_2]$ is the union of all values of μ for which the corresponding acceptance interval is intercepted by the vertical line.

Data x	**μ range**
<1.5	Empty
1.5 – 2.2	Upper limit
>2.2	2-sided

Example:

Param = Temp at centre of Sun

Data = est. flux of solar neutrinos

$\mu \geq 0$

No prior for μ

$\mu_l \leq \mu \leq \mu_u$ at 90% confidence

Frequentist	μ_l and μ_u known, but random μ unknown, but fixed Probability statement about μ_l and μ_u
Bayesian	μ_l and μ_u known, and fixed μ unknown, and random Probability/credible statement about μ

Bayesian versus Frequentism

	Bayesian	Frequentist
Basis of method	Bayes Theorem → Posterior probability distribution	Uses pdf for data, for fixed parameters
Meaning of probability	Degree of belief	Frequentist definition
Prob of parameters?	Yes	Anathema
Needs prior?	Yes	No
Choice of interval?	Yes	Yes (except F+C)
Data considered	Only data you have	….+ other possible data
Likelihood principle?	Yes	No

Bayesian versus Frequentism

	Bayesian	Frequentist
Ensemble of experiment	No	Yes (but often not explicit)
Final statement	Posterior probability distribution	Parameter values → Data is likely
Unphysical/ empty ranges	Excluded by prior	Can occur
Systematics	Integrate over prior	Extend dimensionality of frequentist construction
Coverage	Unimportant	Built-in
Decision making	Yes (uses cost function)	Not useful

Bayesianism versus Frequentism

"Bayesians address the question everyone is interested in, by using assumptions no-one believes"

"Frequentists use impeccable logic to deal with an issue of no interest to anyone"

Approach used at LHC

Recommended to use both Frequentist and Bayesian approaches

If agree, that's good

If disagree, see whether it is just because of different approaches

Tomorrow (last lecture)

Comparing data with 2 hypotheses

H0 = background only (No New Physics)

H1 = background + signal (Exciting New Physics)

Specific example: Discovery of Higgs