Bayes v Frequentism

Return of an old controversy

Louis Lyons
Oxford

Bayes v Frequentism

Upper Limits

Workshops

Comparison of methods

Feldman & Cousins

CLs

Where are we now?
How can text-books not even mention Bayes/Frequentism?

For simplest case \( (m \pm 3\sigma) \)

with no constraint on \( m_{\text{true}} \),

\[ m - k \sigma < m_{\text{true}} < m + k \sigma \]

at some prob

for both

(but different interpretations)

See:
Bob Cousins "Why isn't every physicist a

\[ \checkmark \quad \checkmark \quad \checkmark \]
\[ \checkmark \quad \checkmark \quad \checkmark \]

2/1045"
NEED TO MAKE STATEMENT ABOUT param, GIVEN data

BASIC DIFFERENCE

BAYES:

\[ \text{Prob (param | data)} \]

Anathema to Frequentist

FREQUENTIST

\[ \text{Prob (data | param)} \]

Likelihood Function
1) **BAYES**

\[ P(A \text{ and } B) = P(A|B) \times P(B) = P(B|A) \times P(A) \]

- e.g. \( A = \) event contains 6 quark
  \( B = \) event contains W boson

- or \( A = \) you are in FNAL
  \( B = \) you are at Workshop

Completely uncontroversial.

\[ P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \]
BAYES:

\[ P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \]

Bayes Th:

\[ P(\text{Hypothesis} | \text{Data}) \propto P(\text{Data} | \text{Hypothesis}) \]

\[ \uparrow \]

Posterior

\[ \rightarrow \]

PROBLEMS:

1) \( P(\text{Hyp} \ldots) \)

True or False

"Degree of belief"

Credible interval

2) Prior: What functional form?

Uninformative prior:

Flat? In which

\[ \text{e.g. } m, m^2, \ln m \]

Unimportant if "data overshadows prior"

Important for limits

Subjective or objective prior?

3) Coverage
\[ P(\text{data} \mid \text{theory}) \neq P(\text{theory} \mid \text{data}) \]

Theory = male or female

Data = pregnant or not pregnant

\[ P(\text{pregnant} \mid \text{female}) \approx 2\% \]

BUT

\[ P(\text{female} \mid \text{pregnant}) \gg 2\% \]
\[ P(\text{Data} \mid \text{Theory}) \neq P(\text{Theory} \mid \text{Data}) \]

**HIGGS SEARCH AT CERN**

Is data consistent with S.M.

or with S.M + Higgs?

End Sept: Data not very consistent with S.

\[ \text{Prob (Data (S.M.) < 1\%)} \]

Frequentist statement

Turned by Press into

\[ \text{Prob (S.M. \mid data) < 1\%} \]

and hence

\[ \text{Prob (Higgs \mid data) > 99\%} \]

i.e. "IT IS ALMOST CERTAIN THAT THE HIGGS HAS BEEN SEEN"
Example 1: Is coin fair?

Toss coin: 5 consecutive tails

What is

\[ P(\text{unbiased|data})? \quad \text{i.e. } p = \]

Depends on prior \( (\theta) \)

If coinage biased: prior \( \sim \mathcal{U}(0,1) \)

Example 2: Particle identification

Try to separate \( \pi^+ \) and protons

\[ \text{prob}(p \text{ tag | real } p) = 0.95 \quad \text{prob}(\pi^+ \text{ tag | real } \pi) \]

\[ \text{prob}(\pi^- \text{ tag | real } p) = 0.10 \quad \text{prob}(p \text{ tag | real } \pi) \]

Particle gives proton tag. What is it?

Depends on prior = fraction of protons.

If proton beam, very likely

If general secondary particles, more even

If bare \( \pi \) beam, \( \times \circ \).
\[
\text{Prob (Data | Theory)} \neq \text{Prob (Theory | Data)}
\]

Given that:

1. "Dog \text{d} \text{ has } 50\% \text{ prob of being within } 100 \text{ m of hunter } \text{h}."

is it true that

2. "Hunter \text{h} \text{ has } 50\% \text{ prob of being within } 100 \text{ m of dog } \text{d}?"

Additional information:

"Dog can be at -10 m, hunter cannot be at -10 m."

If dog at -10 m, hunter cannot be within 100 m of dog.
Neyman "confidence interval" avoids false 

Keeps to $P(x | \mu)$

Confidence interval $\mu_1 \to \mu_2$:

$P(\mu_1 \to \mu_2 \text{ contains } \mu) = \alpha$ \quad \text{True for}

Varying intervals fixed from ensemble of samples

Gives range of $\mu$ for which observed val $x_0$ was "likely"

[Contrast Bayes: Degree of belief that $\mu$

is in $\mu_1 \to \mu_2$ is $\alpha$]

If true for all $\mu$ $P < \alpha$ for some $\mu$ "correct" $\alpha$

$P > \alpha$ "undercoverage" "serious"

$P > \alpha$ "overcoverage" "conservative"
CLASSICAL INTERVALS: PROBLEMS

Hard to understand (d'Agostini et al.)
- Arbitrary choice of interval
- Possibility of empty range
  (overcoverage for integer observation e.g. no. of events)
- Nuisance parameters (systematic errors)

ADVANTAGES

- Widely applicable
- Well defined coverage
FIG. 1. A generic confidence belt construction and its use. For each value of $\mu$, one draws a horizontal acceptance interval $[z_1, z_2]$ such that $P(z \in [z_1, z_2] | \mu) = \alpha$. Upon performing an experiment to measure $z$ and obtaining the value $z_0$, one draws the dashed vertical line through $z_0$. The confidence interval $[\mu_1, \mu_2]$ is the union of all values of $\mu$ for which the corresponding acceptance interval is intercepted by the vertical line.
Pour \( \xi \in \mathbb{C} \) et \( \eta \in \mathbb{C} \), on a une valeur \( \xi^2 + \eta^2 = a^2 + b^2 \).
Where do we go? This leaves us...
I. WORKSHOPS

observed events < observed bag

meas om x user / out / the physical bond

big and well known

small or unobserved signal

1. Theories on
CONFIDENCE LIMITS WORKSHOPS

CERN, January 2000

FNAL, March 2000

CERN: Yellow Report 2000-005

+ Website

http://cern-web.cern.ch/CERN/Divisions/P1/Events/CLW/

FNAL: Website

http://conferences.fnal.gov/cl2k/
Workshop on 'Confidence Limits'
17-18 January, 2000
CERN Council Chamber

This workshop will be devoted to the problem of setting confidence limits in difficult cases: small or unobserved signal, background larger than signal, background not well known, and measurements near a physical boundary. Among the many examples in HEP are: the Higgs mass, accelerator searches for neutrino oscillations; $B_s$-decays, SDSY; compositeness; neutrino masses; and dark matter. Several different methods and of the market - the GL methods used by the LEP Higgs searches: Bayesian, Feldman-Cousins and modifications thereof, etc.

Distinguished invited speakers will present the methods currently being proposed, and all participants are encouraged to submit abstracts for short contributed papers, of which a limited number will be selected for presentation. Proceedings will include summaries of all papers presented, as well as discussions.

<table>
<thead>
<tr>
<th>Invited Speakers</th>
<th>Co-convenors</th>
<th>Local Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter Clifford / Oxford</td>
<td>Fred James / CERN</td>
<td>Louis Lyons / Oxford</td>
</tr>
<tr>
<td>Bob Cousins / UCLA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giulio D'Agostini / Roma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carlo Giunti / Torino</td>
<td>Fred James</td>
<td></td>
</tr>
<tr>
<td>Fred James / CERN</td>
<td>Yves Perrin</td>
<td></td>
</tr>
<tr>
<td>Harrison Prosper / Florida State</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alex Read / Oslo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michael Woodroofe / Michigan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guenter Zech / Siegen</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Open to all who register. There is no conference fee. All participants will be expected to be familiar with the methods under discussion by having read the publications on the list of 'required reading'. Useful background material will also be presented in the Academic Training lectures of F. James at CERN, 10-14 January, where confidence limits will be discussed from different points of view.

If you plan to participate, please register by filling the registration form on the GL Workshop website at:

Proceedings
We will produce proceedings which will contain summaries of all the invited talks and submitted papers, as well as the important points brought up during the discussions, to be published as a CERN Yellow Report.
Le CERN affronte le nouveau millénaire avec confiance

Le Conseil du CERN a ouvert la séance en formulant des remarques sur l'état du CERN à l'ordre d'un nouveau millénaire. "Depuis sa création, le CERN a apporté des contributions remarquables au monde de la science, de la technologie et de l'éducation. Aujourd'hui, à l'aube du nouveau millénaire, je crois que le meilleur est encore à venir. Le CERN est bien préparé à relever les défis de la mondialisation des institutions de recherche. Le trait dominant des prochaines décennies, à l'évidence aussi dans les domaines de la physique des grandes énergies et de l'informatique, sera la concurrence et la coopération mondiales au sein d'alliances stratégiques. Pour réussir dans ce contexte, il

CERN confronts the New Millennium with Confidence

The CERN Council, where the representatives of the Member States of the Organization decide on scientific programmes and financial resources, held its 114th session on 17 December under the chairmanship of Dr. Hans C. Eschelbacher (DE).
research material.

90% had read > 50% of material.

Write a paper about limits.

This expert discovered nothing, so we...
Panel Discussion

Be Ming (P.E.)

Dezhi T.C.
Monday am 1  James Choosing foundat
am 2  Proser  Bayes
         Berger

pm 1  Linnemann Bayes
      Kruse       Searches
      Maeshima
      James       Reply to Bayes

pm 2  Murray CLs
       Raja
       Narzuy

Tuesday am 1  Cousins F+C
             Rolke
             Messier
             Kafka

am 2  Schee CDMS
       Yelin
       Roe

pm  Feldman Methodology
      Murray
      Greenlee
      Cowan     Combining results

CONCLUSIONS
Bayesian TALKS

d'Agostini

Probability for Bayesians

Barlow's book

Very critical of Classical approach

e.g. What does limit on $M_H$ mean

Recently advocates L ratio for testing hypothesis

Used for Higgs limit [d'A + Degrassi]

\[
\frac{P(s | n_{\text{obs}}, b)}{P(s_2 | n_{\text{obs}}, b)} = \frac{P(n_{\text{obs}} | s_1, b)}{P(n_{\text{obs}} | s_2, b)} \times \frac{\text{Prior}(s_1)}{\text{Prior}(s_2)}
\]

Rewritten as

\[
\frac{P(s | n_{\text{obs}}, b)}{\text{Prior}(s)} \times \frac{P(s = 0 | n_{\text{obs}}, b)}{\text{Prior}(s = 0)} = \frac{P(n_{\text{obs}} | s, b)}{P(n_{\text{obs}} | s = 0, b)}
\]

Evidence of signal

\[\begin{array}{c}
\text{No discrim.} \\
\text{Signal coloured}
\end{array}\]

\[\begin{array}{c}
\text{No discrim.} \\
\\text{Rejected}
\end{array}\]
In cancer treatment, Biocomial therapy is used.

\[ \text{Example: } e = 97\% \]
\[ V_{\text{final}} = V_{\text{physical}} + V_{\text{error}} \]

\[ V_{\text{final}} = \text{physical} \]
**LIMITS**

Observe \( \{ \text{zero events} \} \) \( \leq 0 \) \( \rightarrow \) set limit rather than give \( \nu \) (but see Feldman Cousins)

Limit less useful than \( \uparrow \) value \( \pm \) error \( \uparrow \) 2 numbers

Cannot combine limits
Can combine estimates

Assume Gaussian dist without physical region prob

One-sided 90\% 95\% 99\%

- 1.28 \( \approx \) 1.645 \( \approx \) 2.335

Bayesian \( p_0 < \bar{p}_{\text{meas}} + 1.28 \)

What to do if \( \{ \text{large} \} \) is unphysical? \( \bar{p}_{\text{meas}} \)
UNIFIED APPROACH TO CLASSICAL STATISTICAL ANALYSIS OF SMALL SAS

Gary Feldman & Bob Cousins

HUTP-97-4096


SOLVES
1) PROBLEM OF INCORRECT CONFIDENCE INTERVALS FOR "FLIP-FLOP" APPROACH
2) NON-PHYSICAL CONFIDENCE INTERVALS
3) AVOIDS OVERCOVERAGE

EXAMPLES
4) GAUSSIAN WITH PHYSICAL BOUND
5) POISSON COUNTING, WITH B&L
6) OSCILLATIONS
BAYESIAN v CLASSICAL (FREQUENTIST)

CLASSICAL CAN GIVE UNPHYSICAL ANSWER
  e.g. \( m^2 < -200 \)

MOTIVATED BAYESIAN APPROACH (see e.g. PDC)

HERE, FREEDOM IN CLASSICAL APPROACH EXPLOITED \( \Rightarrow \)

LIMITS ALWAYS PHYSICAL

\[ \therefore \text{NO NEED FOR BAYES}\]

ALSO DE COUPLES \( \{ \text{INTERVAL LIMITS} \} \) \text{GOODNESS OF FIT} \]
90% classical interval for Gaussian
\[ \sigma = 1 \]
\[ \mu \geq 0 \]
\[ z_90 \]
\[ \sigma \]

FIG. 3. Standard confidence belt for 90% C.I. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

\[ x_{\text{obs}} = 3 \quad \text{Two sided limit} \]
\[ x_{\text{obs}} = 1 \quad \text{Upper limit} \]
\[ x_{\text{obs}} = -2 \quad \text{No region for } \mu \]
FIG. 10. Plot of our 90% confidence intervals for mean of a Gaussian constrained to non-negative, described in the text.
FIG. 2. Standard confidence belt for 90% C.L. upper limits for the mean of a Gaussian, in units of the rms deviation. The second line in the belt is at $z = +\infty$. 
FIG. 4. Plot of confidence belts implicitly used for 90% C.L. confidence intervals (vertical intervals between the belts) quoted by flip-flopping Physicist X, described in the text. They are not valid confidence belts, since they can cover the true value at a frequency less than the stated confidence level. For $1.36 < \mu < 4.28$, the coverage (probability contained in the horizontal acceptance interval) is 85%.

Not good to let $x_{\text{obs}}$ determine how result will be presented.

F-C goes smoothly from 1-sided $\Rightarrow$ 2-sided.
**ORDERING PRINCIPLE**

Choice of horizon band of Fig 1 is ARBITRARY.

**ORDERING**

At any \( M \), choose \( x \) values according to:

**ORDER**

\[
R = P(x) / P(\max P)
\]

**VALUE** of \( M \) in range \( \Rightarrow \max P \)

**E.G.** Gaussian case:

\[
P(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)
\]

Then add in \( x \) values (for \( \max + 1 \) \( R \))

\[
\int_{x_1}^{x_2} P(x) \, dx = \alpha
\]

\( (x_1, x_2 \) numerically)  

**Fig 10:** Confidence belt (cf. Fig 9)

N.B.

For \( x < 1.28 \), conf. belt \( \Rightarrow \) upper limit

\( x = 0 \), upper limit is 1.64 (not 1.28)

Choose limit caused by regarding under-recovery.
FIG. 6. Standard confidence belt for 90\% C.L. central confidence intervals, for unknown Poisson signal mean $\mu$ in the presence of Poisson background with known mean $b = 3.0$. 
FIG. 7. Confidence belt based on our ordering principle, for 90% C.L. confidence interval unknown Poisson signal mean $\mu$ in the presence of Poisson background with known mean $b = 3.0$. 
BAYES

Easy to understand

Physical intervals

Needs prior

Hard to combine

Coverage
STANDARD  FREQUENTIST

COVERAGE

HARD TO UNDERSTAND

SMALL \( \bar{Z} \) INTERVALS

EMPTY

DIFFERENT UPPER LIMITS
UNIFIED
LIFTS OFF FROM ZERO Too Soon

No Flip-Flop

Coverage

Hard To Understand

Decreasing Limit as BGD > Obs

Computing & Time
Try to overcome problem of limit $b$ as $b \geq b_{0}$.

F & C were aware, but decided to live with it.

Defined "Sensitivity".

Roe & Woodroofe:
Condition on $b_{0} \leq b_{obs}$

Bouchez: One side Bayes.
Other side = frequentist coverage.

Mandelker & Shultz

Harrison Prosser

"Freedom to do what you want."
COMBINING LIMITS

COMBINING MEASUREMENTS is "EASY"

\[ x_i \pm \sigma_i \Rightarrow \bar{x} = \frac{\sum x_i / \sigma_i^2}{\sum 1 / \sigma_i^2} \]

\[ \frac{1}{\sigma^2} = \sum \frac{1}{\sigma_i^2} \]

WHY IS IT IMPOSSIBLE TO COMBINE LIMITS \( L_i \)?

1) \( x_i \pm \sigma_i \) = 2 numbers

\( L_i \) = 1 number: Not enough information

2) Measurement: \( \sigma_i^2 \) = variance.

Limit: No information about distribution.
SUMMARY

Comments:
- Too polite
- Too theoretical

Statistics:
- Don't reinvent the wheel
- Not been after the devil-attraction
- Useful to explain what we are doing

Lots of detailed, improved understanding

Dispel misconceptions

Cousin's 10 points

Experts found it useful to interact
- Going away to think
- Problems to attack

1) Nuisance parameters? in Classical approach

2) Combining results
   - e.g. What do you need for F-C?
   - [Fred James: What do you want it for?]
IX. What Might We Agree On?

Maybe quite a bit!

1. First of all, *civility*. Bohr and Einstein make better *role models* than Lindley and others in the statistics community.

2. $P(\text{hypothesis} \mid \text{data})$ cannot be calculated without a prior.

3. The likelihood function is not a pdf in the unknown parameters. "Integrating the likelihood function" is not a concept in either Bayesian or classical statistics. Uniform priors should be explicitly stated, not hidden.

4. Answers based on "uniform priors" depend on the metric in which the prior is uniform.

5. The problem of upper limits cannot be considered solved without having a consistent method for two-sided intervals.

6. Bayesian intervals typically do not have frequentist coverage.

7. Publishing enough info to reconstruct an approximate likelihood function should be strongly encouraged.

8. Our usual gof tests do not exist in Bayesian statistics: must be reformulated as extending the space of $P$.

9. The confidence interval construction does not use a prior. It uses $P(\text{data} \mid \text{theory})$, which requires the ensemble to be specified. Priors enter when going from $P(\text{data} \mid \text{theory})$ to $P(\text{theory} \mid \text{data})$, which confidence intervals do not do.

10. Regardless of your opinion about priors, a subjective utility function is needed to make a decision, so any argument for totally objective decisions is highly suspicious.
Clarification of many specific points
No "Grand Unified Method" yet

Review for Rev Mod Phys