THE POWER OF CONFIDENCE INTERVALS

Carlo Giunti

INFN, Sezione di Torino, and

Dipartimento di Fisica Teorica, Università di Torino

giunti@to.infn.it

C.G, Marco Laveder, NIMA 480 (2002) 763, hep-ex/0011069

$\boxed{\text{POWER} \Longleftrightarrow \text{RELIABILITY}}$

Advanced Statistical Techniques in Particle Physics

University of Durham (UK)

18–22 March 2002

RELIABILITY of Confidence Intervals

a reliable Confidence Interval carries real physical information on the parameter under investigation

reliable = physically significant

experimentalists should be interested in reporting reliable Confidence Intervals that can be used safely by other scientists (what is the use of compilations of Confidence Intervals as that in the Review of Particle Properties if they are not reliable?)

examples of unreliable confidence intervals:

☹ empty Confidence Interval (no info, useless)

☹ Upper Limit much below sensitivity (misleading info)

such intervals are statistically correct!

but we are not interested in mathematical entities

we are interested in physics!

unreliable Confidence Intervals are harmful for physics

still problem in understanding reliability?

then forget the Unified Approach, use Central Intervals and give empty Confidence Intervals when they appear

Main Property of Confidence Intervals

COVERAGE

a $100(1-\alpha)\%$ Confidence Interval belongs to a set of intervals that cover the true value of the parameter μ with probability $1-\alpha$.

COVERAGE is not enough!

many sets of intervals constructed with different methods have COVERAGE but different POWER – Neyman's construction of acceptance intervals based on hypothesis testing

COVERAGE $\Longleftrightarrow$ type I error

reject the null hypothesis μ_0 when it is true

$\alpha =$ probability of type I error ("size" of test)

POWER $\Longleftrightarrow$ type II error

accept μ_0 when μ_1 is true

$\beta =$ probability of type II error

POWER $=$ probability to reject μ_0 when μ_1 is true

$$\pi = 1 - \beta$$

best choice: Most Powerful Test

PROBLEM: $\mu_1 \neq \mu_0$ is a composite hypothesis

$$\mu_1 \neq \mu_0 \Longrightarrow \pi(\mu_1)$$

Uniformly Most Powerful Test?

in general a Uniformly Most Powerful Test does not exist if the alternative hypothesis is two-sided ($\mu_1 < \mu_0$ and $\mu_1 > \mu_0$) unless the class of tests is restricted with additional requirements

Unbiased Tests

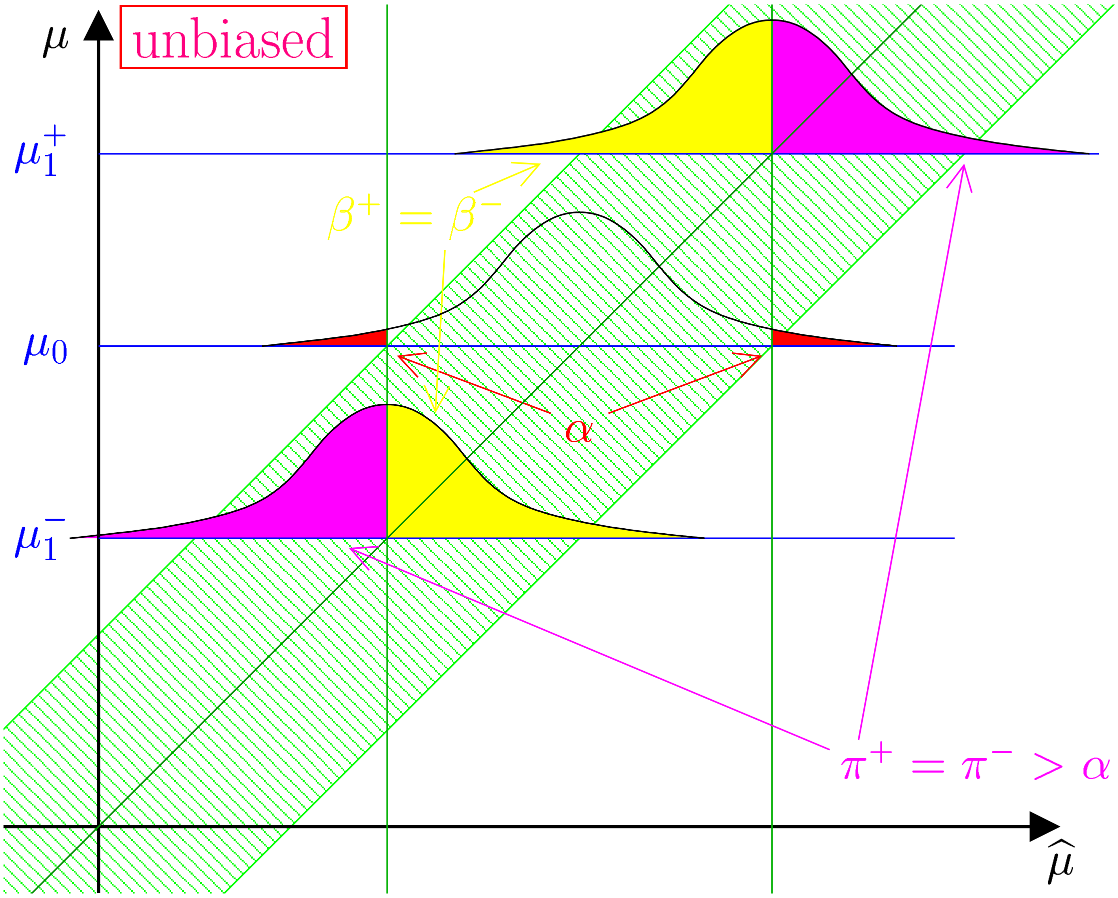
the probability of rejecting μ_0 when it is false is at least as large as the probability of rejecting μ_0 when it is true.

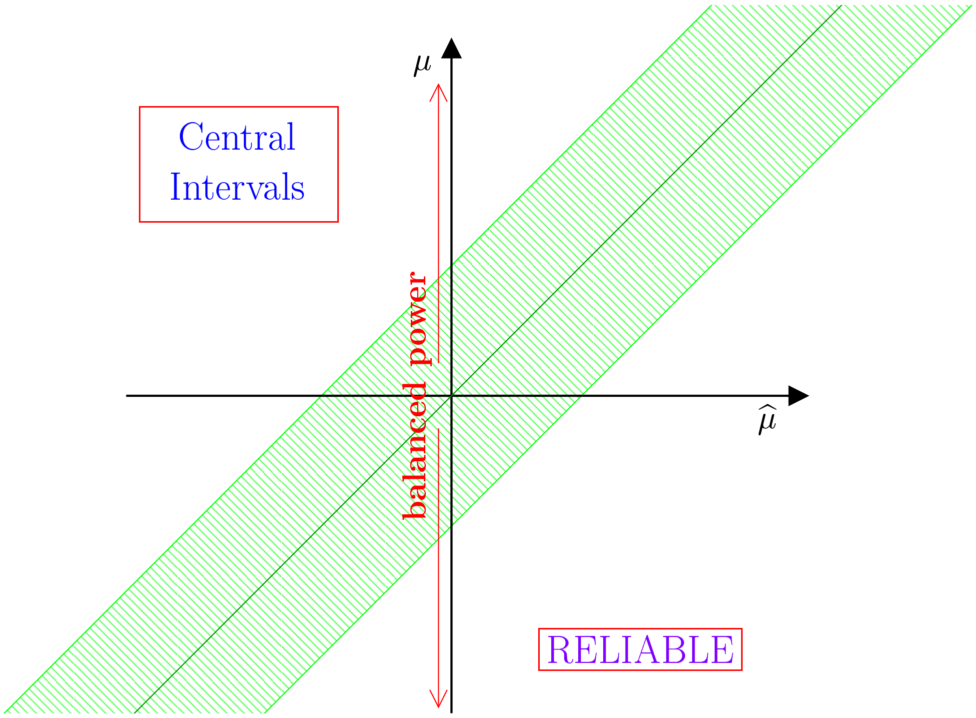
$$\inf_{\mu_1} \pi(\mu_1) \geq \alpha$$

the equal-tail test used in the Central Intervals method is Uniformly Most Powerful Unbiased for distributions belonging to the exponential family (Gauss, Poisson, ...)

Central Intervals

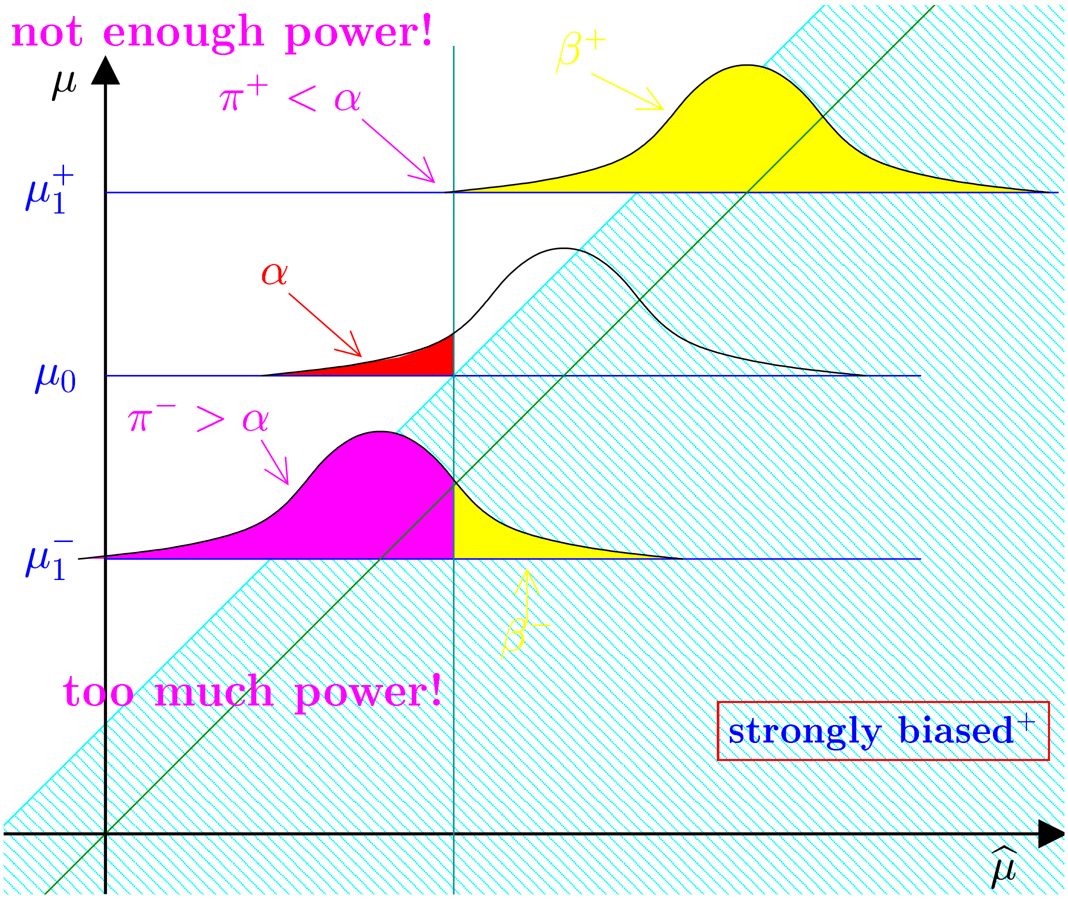
unbounded μ – estimator $\widehat{\mu}$ with Gaussian distribution

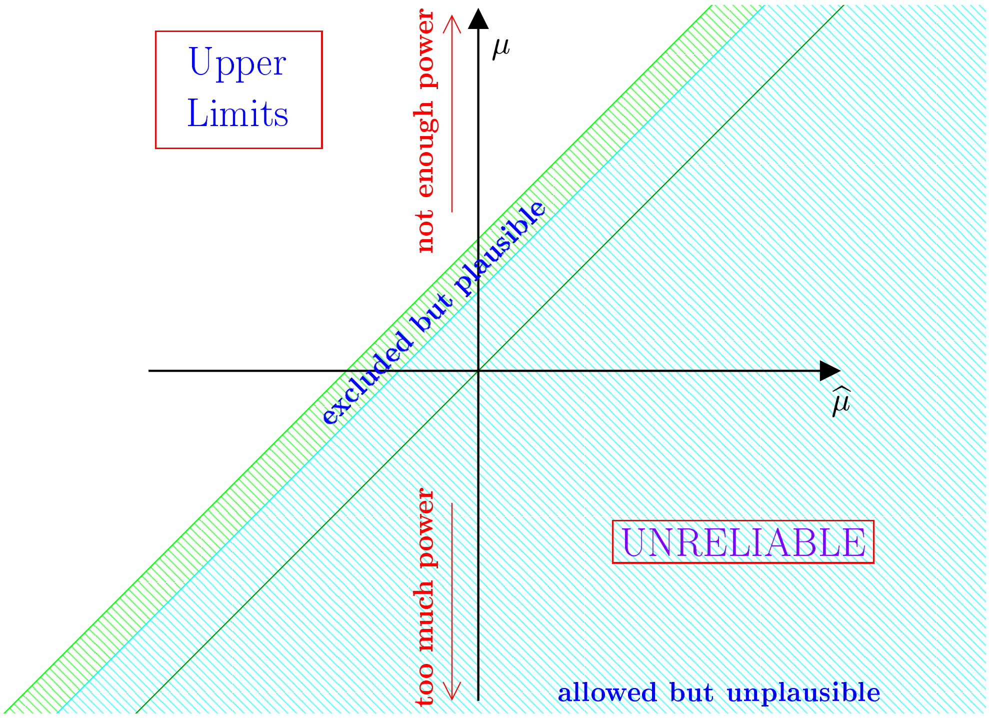




unrealistic example: Upper Limits

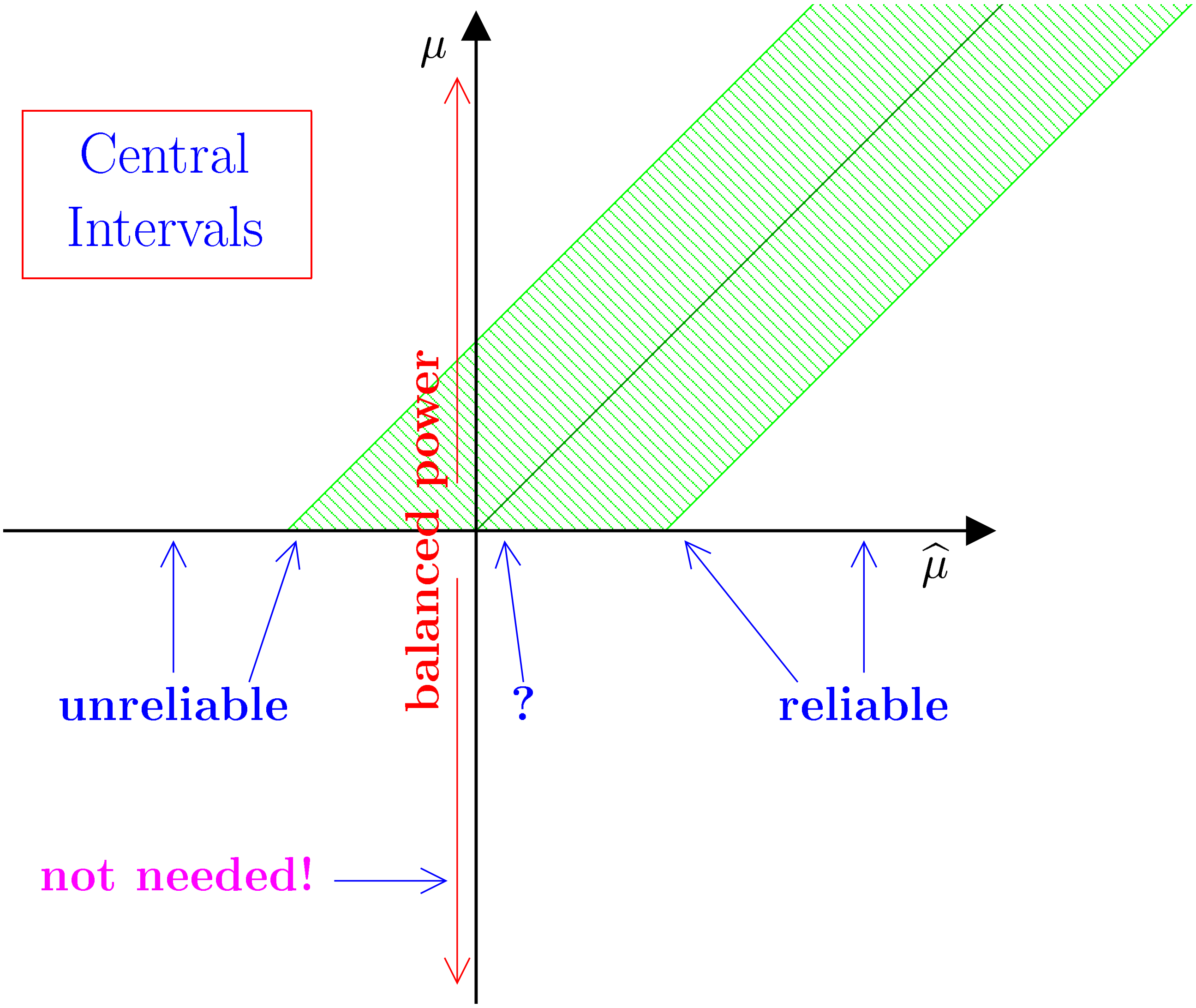
unbounded μ – estimator $\widehat{\mu}$ with Gaussian distribution





Central Intervals

bounded $\mu \geq 0$ – estimator $\widehat{\mu}$ with Gaussian distribution



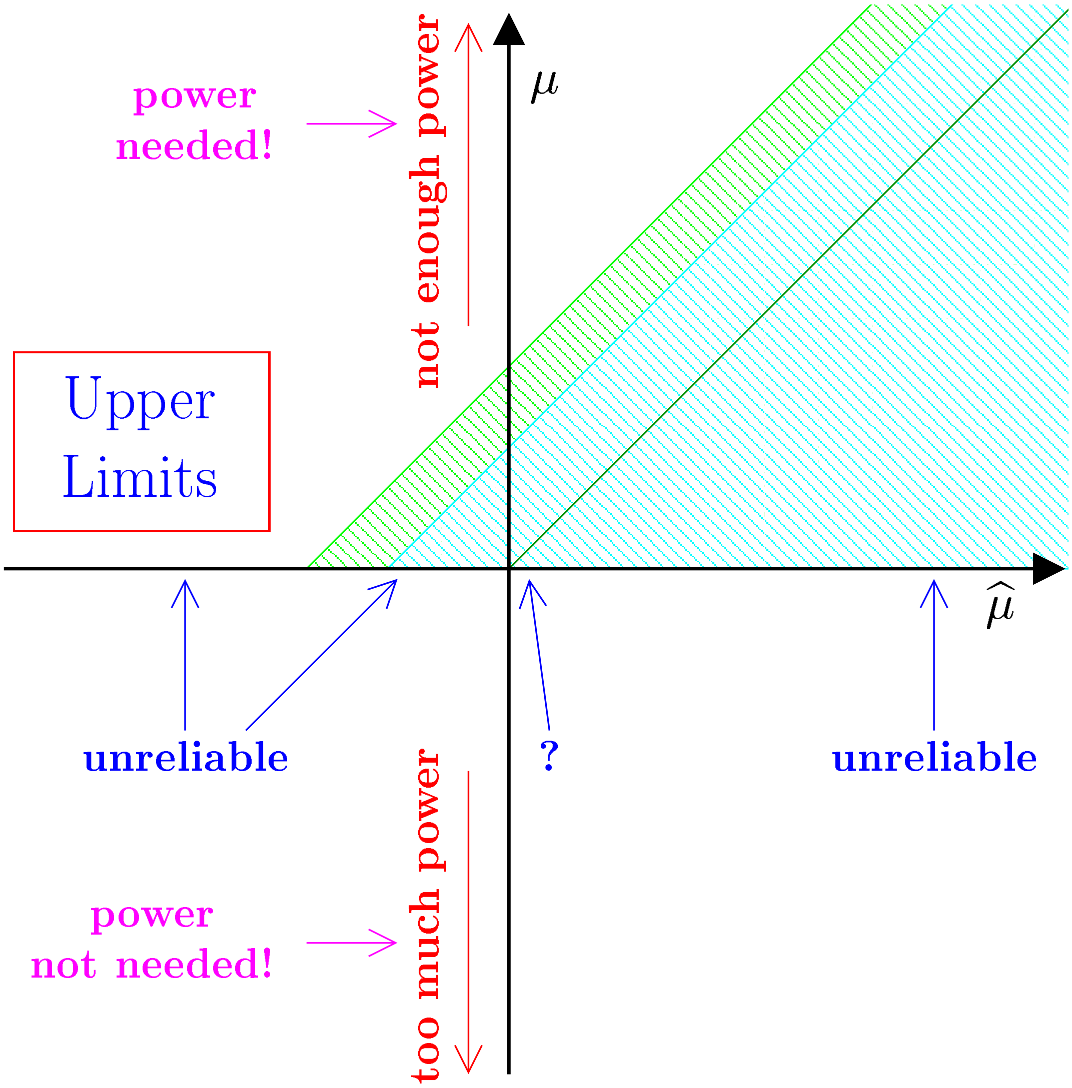
power to test $\mu_1 < \mu_0$ is not needed for small μ_0 !

more power to test $\mu_1 > \mu_0$ is needed for small μ_0 !

NEED METHOD RELIABLE FOR ALL $\widehat{\mu}$!

Upper Limits

bounded $\mu \geq 0$ – estimator $\widehat{\mu}$ with Gaussian distribution

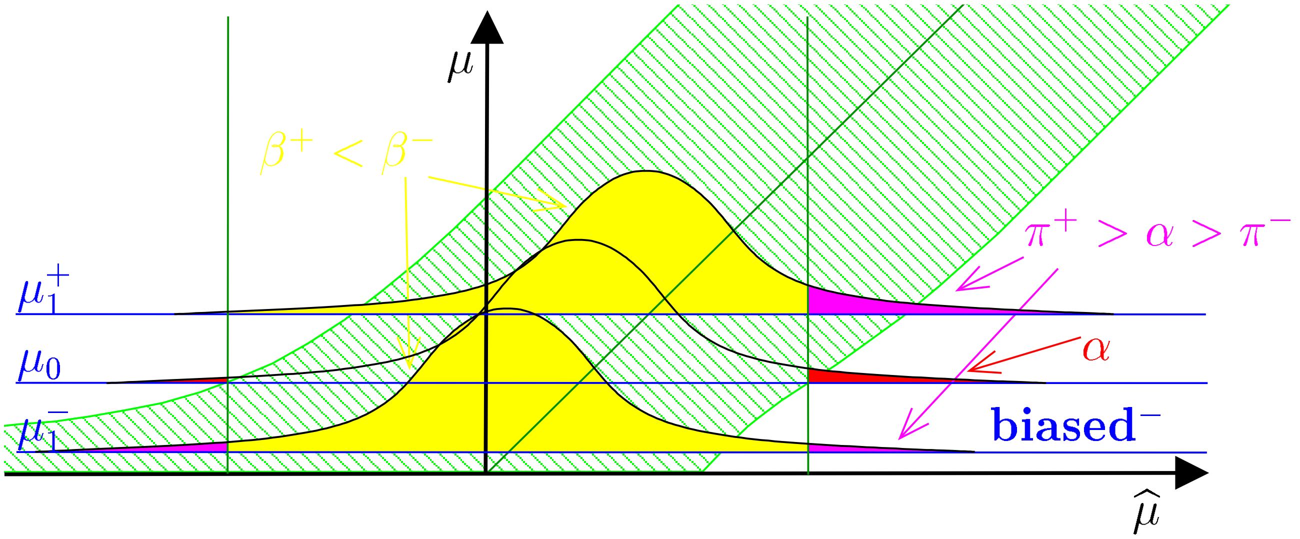


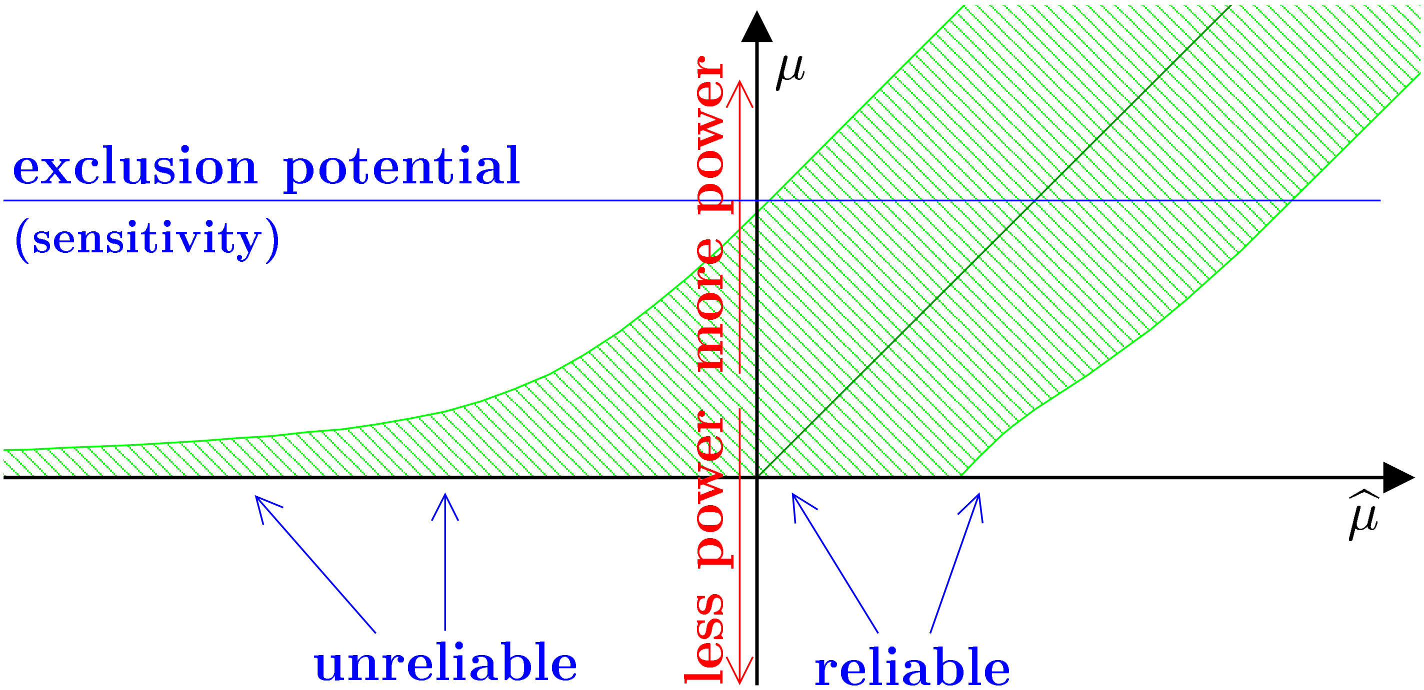
WORSE THAN CENTRAL INTERVALS !

too much power where not needed !

Unified Approach

[G.J. Feldman and R.D. Cousins, Phys. Rev. D 57 (2998) 3873, physics/9711021]





real breakthrough !

more power to test upper deviations for small μ

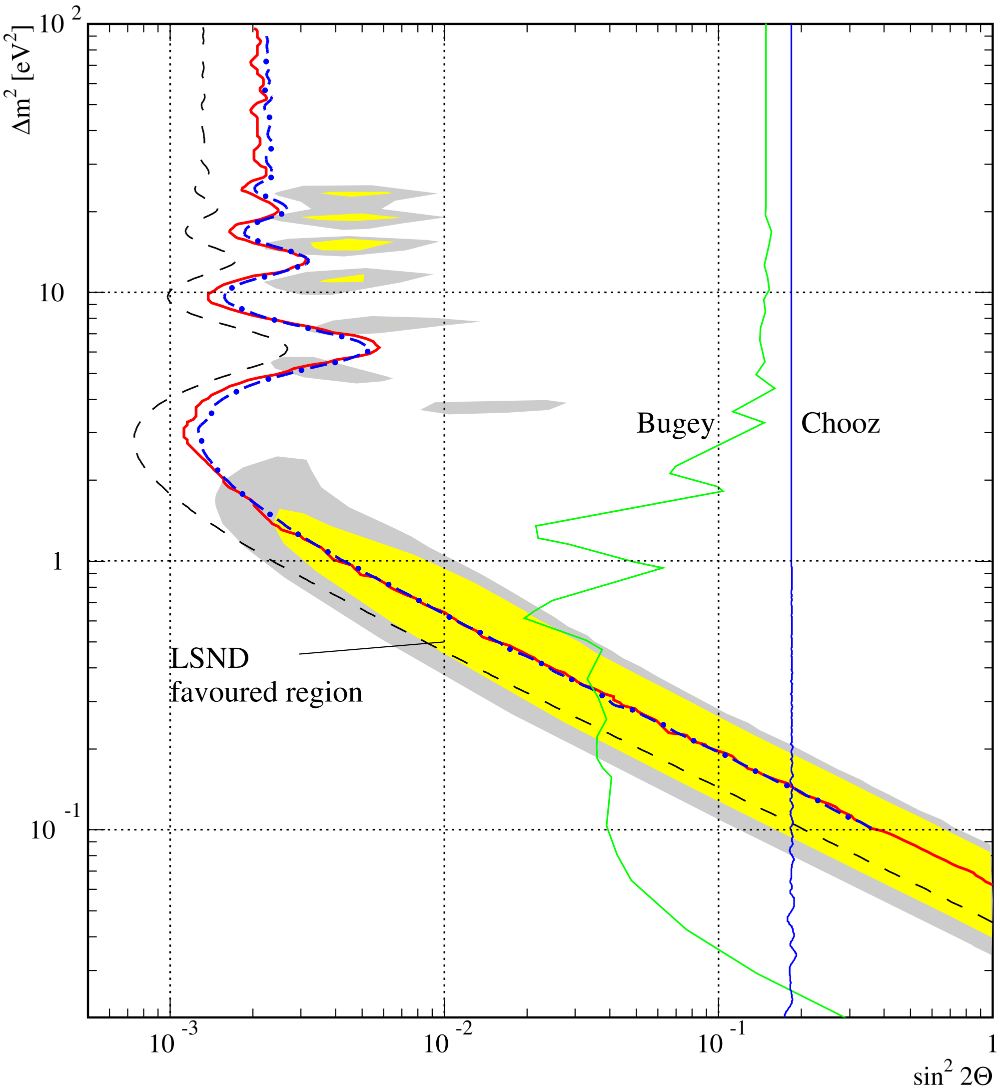
right direction but not enough !

REAL EXAMPLE

KARMEN experiment: $\bar{\nu}_\mu \to \bar{\nu}_e$ oscillation search

GREAT EXPERIMENT!

hep-ex/9908043, Moriond 1999

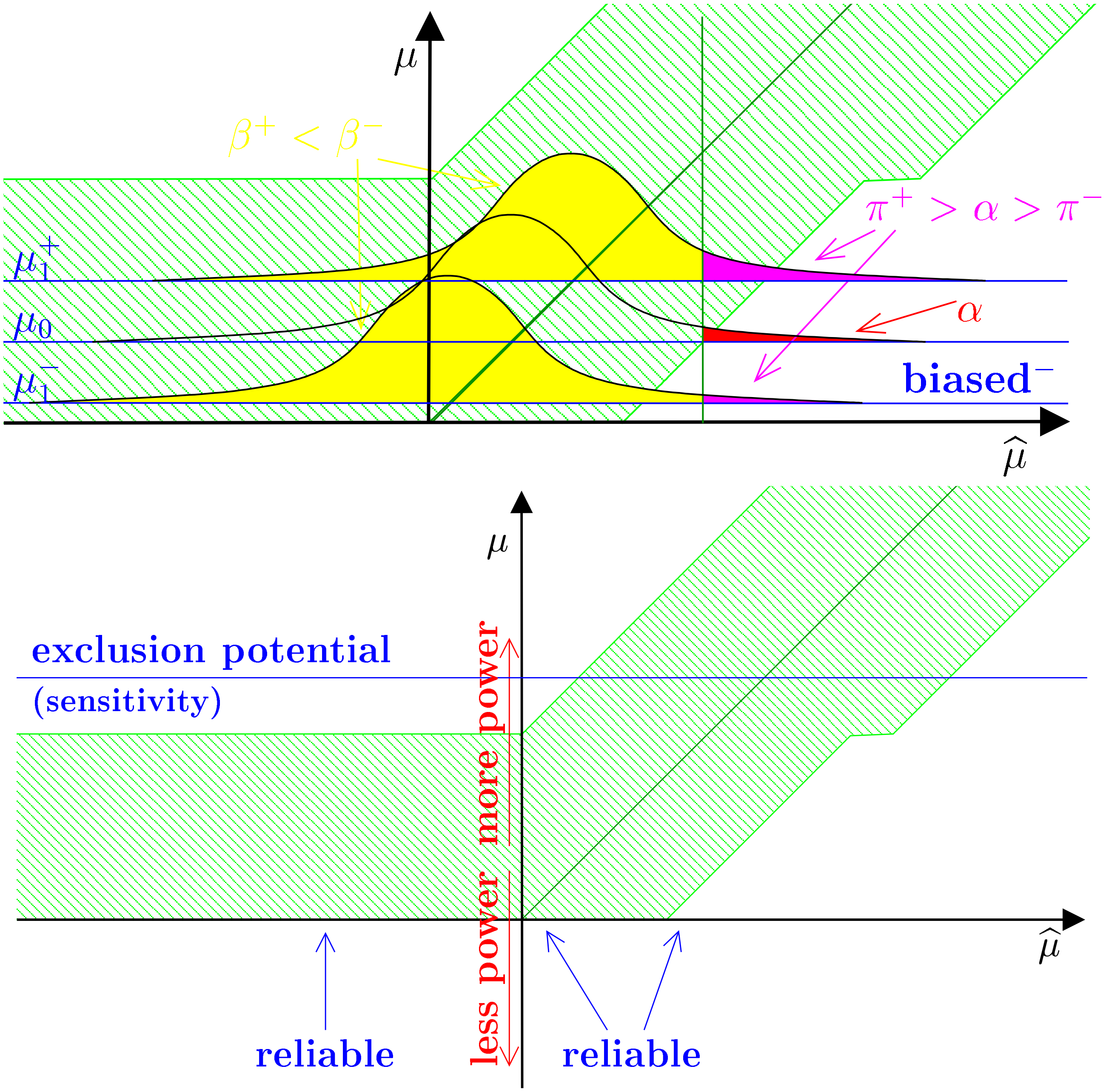


KARMEN 2 90% confidence limits according to the Unified Approach compared to other experiments: The full line is the 90% C.L. of the data presented here, the dotted line the corresponding sensitivity and the dashed line the 90% C.L. derived from the Feb. 97 – Apr. 98 data.

Maximum Likelihood Estimator

[S. Ciampolillo, Il Nuovo Cimento A 111 (1998) 1415]

[M. Mandelkern and J. Schultz, J. Math. Phys. 41 (2000) 5701, hep-ex/9910041]



OPTIMAL POWER !

COMMENTS

- the boundary $\mu \geq 0$ introduces an asymmetry in the importance of the $\mu_1 > \mu_0$ and $\mu_1 < \mu_0$ alternatives
- testing the alternatives $\mu_1 > \mu_0$ is much more important (and physically more meaningful) than testing the alternatives $\mu_1 < \mu_0 \Longrightarrow$ bias is desirable
- more power in testing $\mu_1 > \mu_0 \Longrightarrow$ values of μ smaller than the true one are less likely to be accepted $\Longrightarrow$ the experiment is more sensitive to a possible signal
- less power in testing $\mu_1 < \mu_0 \Longrightarrow$ less stringent upper limits $\Longrightarrow$ more reliable

CONCLUSIONS

We have considered a Gaussian distribution with $\mu \geq 0$. Similar considerations apply to a Poisson distribution with background.

POWER $\Longleftrightarrow$ RELIABILITY

Maximum Likelihood Estimator
better than
Unified Approach
better than
Central Intervals
better than
Upper Limits